UNIVERSITY OF SOUTHAMPTON

FACULTY OF ENGINEERING, SCIENCE AND MATHEMATICS

Optoelectronics Research Centre

An Electrically Driven 8-14 um Band Solid-State Modulator

by

Chong Yew Lee

Thesis for the degree of Doctor of Philosophy

December 2008

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF ENGINEERING, SCIENCE AND MATHEMATICS

OPTOELECTRONICS RESEARCH CENTRE

Doctor of Philosophy

AN ELECTRICALLY DRIVEN 8-14 um BAND SOLID-STATE MODULATOR

by Chong Yew Lee

The need for a solid state broadband modulator that is efficient and compact, to replace the functionality of mechanical choppers used to modulate pyroelectric detector signals, has existed for some time. Systems where mechanical solutions are undesirable, especially when there are space and power constraints, and also devices sensitive to vibrations will benefit from the use of solid state modulators.

Requirements that must be satisfied for solid state modulators to be feasible include good transmission, adequate depth of modulation, low power consumption, insensitivity to polarization, a wide spectral bandwidth and easy integration into optical systems with high numerical aperture. The research in this project has led to the first successful fabrication of an electrically operated solid state modulator for the 8 to 14 micron region, which uses intervalence carrier transitions to induce modulations. An average 60% reduction in transmission across a 5 mm aperture was achieved at less than 3W of power dissipation. Temporal diagnostics using a carbon dioxide laser probing have shown that the modulator is capable of operating at frequencies up to at least 200Hz, which is in agreement with simulation results.

Doped regions on the modulator were also successfully produced using solid state epitaxial regrowth, offering a cheaper alternative to ion implantation.

The future use of thicker wafers to fabricate modulators would result in a greater depth of modulation. Because the present design of the modulator suffers from uneven current flow through its electrodes due to parasitic resistances, the use of current mirrors is proposed to force current through different regions of the electrode to improve its performance.

Contents

List of Figures

List of Tables

DECLARATION OF AUTHORSHIP

I, Chong Yew Lee, declare that the thesis entitled: **An Electrically Driven 8-14 um Band Solid-State Modulator** and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research. I confirm that:

- this work was done wholly or mainly while in candidature for a research degree at this University;
- where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- where I have consulted the published work of others, this is always clearly attributed;
- where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- none of this work has been published before submission.

Signed:

Date: DECEMBER 2008

Acknowledgement

I praise and thank my Lord and my God, Jesus Christ, for saving mankind from the wages of sin, which is death, by dying on the cross as an atoning sacrifice for our sins; thus freeing us from the throes of death and giving us the hope of an eternal life with God the Father.

I thank Harvey for giving me the opportunity to complete a PhD in Optoelectronics. I have lost count of the number of times I caused his blood pressure rise to dangerous levels. How he managed to maintain his composure every single time is admirable. Maybe that prepared him for the task as the head of school? His passion for science, general guidance, and sharing about life, the universe and everything (most of which involve diving, flying, climbing and jumping off planes) enriched my experience as a PhD student.

My special thanks to Michael Pollard who thought me how to be creative in solving problems. He gave me the courage to improvise and the enthusiasm to "just do it and see what happens". Well Mike, I got me PhD thanks to you. Happy days.

I thank my parents and my sister and the members of the Chinese Christian Church for their support and encouragement throughout my PhD. They gave me the strength to persevere!

Finally, I thank all the technicians (Ed, Mark Lessey, Mark Long, Eric, Ken, Dave and Neil Sessions); colleagues (Jerry, Mike, Paul, Arun, Everado, and Andrew-ski); and especially to the support staff at the ORC: Eve Smith, Eleanor Tarbox and Angie for their encouragement, advice and help. This project would not have been possible without their assistance!

Chapter 1

Introduction

Prior to the advent of microbolometers, the cost effective method for producing thermal imagers is to use pyroelectric detectors. Pyroelectric detectors belong to the class of thermal detectors, which convert the radiation energy of photons into heat. It is relatively cheap to produce, and has a spectral response range that is only limited by the window material used. However, it is less sensitive compared to photon detectors and requires a mechanical chopper blade to modulate incoming IR radiation (pyroelectric detectors only respond to changes to light). Because mechanical choppers suffer from the drawbacks of size and introduces vibrations to the system, a need has existed for solid state modulators (SSM) to replace the functionality of mechanical choppers. However, the usefulness of a SSM as an optical chopper for *thermal imaging systems* has diminished significantly with the advent of microbolometers[1], which became the choice of detector for producing low to mid range thermal imaging systems[4]. Nevertheless, other applications which uses the pyroelectric detector such as gas sensing and human body detection (e. g., intrusion alarm device) would still benefit from the use of the SSM [6].

Requirements that must be satisfied for SSM to be feasible include good transmission, adequate depth of modulation, low power consumption, insensitivity to polarization, a wide spectral bandwidth and easy integration into optical systems with high numerical aperture. These requirements prevent the use of alternate modulation schemes such as acousto-optic, electro-optic and plasma induced free carrier absorption for modulating light between 8 and 14 μm [2]. However, a novel method for modulating

[1]Microbolometer arrays are preferred because they can be manufactured relatively cheaply using standard silicon wafer fabrication technology, have lower noise equivalent temperature difference (NETD), and are, in general, faster compared to pyroelectric detectors [1, 3, 5].

IR radiation at 8-14 μm that fulfils most of the desiderata for a good SSM has been developed [2]. It is made possible using germanium, which utilizes inter valence band carrier transitions and free carrier absorption to induce IR attenuation (section 2.2.2). A depth of modulation of up to 95.4% was achieved by exciting carriers in germanium using 980 nm laser diodes. Nevertheless, there was a need to further optimize the device in terms of efficiency and design as the use of laser diodes makes the modulator inefficient, costly and bulky.

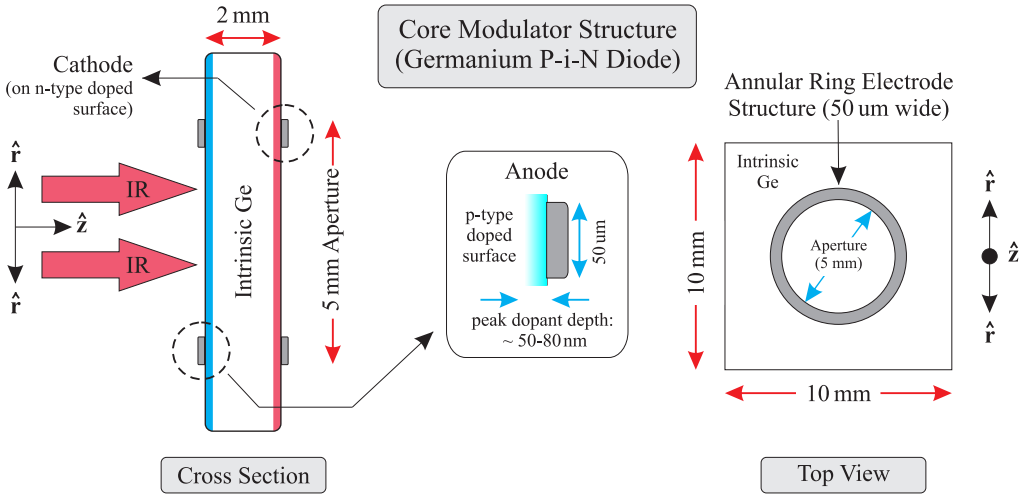


Figure 1.1: The structure of an electrically operated germanium modulator.

A more efficient way of introducing carriers into germanium is through electrical injection. This can be realized by constructing a germanium pin diode structure as shown in figure 1.1. As carriers are injected into the modulator during forward bias, a current flow is established between the electrodes. This gives rise to a gradient in carrier density, which causes the carriers to diffuse laterally in the $\hat{r}$ direction to flood the substrate with IR absorbing carriers. The size of the aperture of the modulator is defined by the diameter of the electrode.

The work in this thesis has led to the first demonstration of an electrically driven, $8-14\,\mu$m band solid state modulator. However, the best performing device only produced an average 60% depth of modulation across the aperture when measured using a FTIR spectrometer. This is caused by uneven current flow between the electrodes as a result of parasitic resistance at the contacts. The use of segmented electrodes, each connected to a different constant current supply should eliminate this problem.

1.1 The Outline of This Thesis

The thesis is organised as follows: chapter 2 outlines the mechanisms that are responsible for inducing IR attenuation in germanium between $8 - 14\,\mu$m. In chapters 3 and 4, analytical equations that describe the electrical properties of germanium are reviewed. Chapter 5 presents the simulation results on the predicted behaviour of the modulator that make use of the physical models described in chapters 3 and 4. Methods to optimize the performance of the modulator are also detailed. The entire procedure to produce the modulator using planar processing technology is covered in chapter 6. Results from the bulk of experiments conducted are presented in chapter 7. These include carrier lifetime measurements, the total depth of modulation, and the electrical characterization of the modulator. Finally, the conclusion of this work along with discussions on future developments to improve the performance of the modulator are given in Chapter 8.

Bibliography

[1] Microbolometers offer high resolution at room temperature. *Opto and Laser Europe*, May 2001.

[2] P. D. Fairley. *Novel Solid State Modulator for the Infrared: The Germanium Chopper*. PhD thesis, University of Southampton, 2000.

[3] V. N. Leonov and D. P. Butler. Two-color thermal detector with thermal chopping for infrared focal-plane arrays. *Applied Optics*, 40(16):2601–2610, 2001.

[4] D. Ostrower. Optical thermal imaging—replacing microbolometer technology and achieving universal deployment. *The Advanced Semiconductor Magazine*, 19(6):24–27, 2006.

[5] J. L. Tissot, C. Trouilleau, B. Fieque, A. Crastes, and O. Legras. Uncooled microbolometer detector: Recent developments at ulis. *Opto-Electronics Review*, 14(1):25–32, 2006.

[6] www.hamamatsu.com. Characteristics and use of infrared detectors. *Technical Information*, 2007.

Chapter 2

Absorption Process in Germanium

IR absorption mechanisms present in germanium (Ge) between 8 to 14 μm band are reviewed in this chapter. Absorptions within this region are mainly caused by the electronic transition of carriers in Ge and are best elucidated with the help of the energy-momentum (E-k) diagram. Therefore, an overview of the E-k diagram for germanium first given in order to aid discussions on IR absorptions in Ge. (Materials presented in this chapter are gathered from various publications on semiconductor physics.)[1, 14, 17–20]

2.1 The Electronic Band Structure (E-k Diagram)

In crystalline germanium, atoms are held together by covalent bonds. When the bond breaks, it behaves as if it were a *mobile* positive charge ($q = 1.6 \times 10^{-19}$ C). This positive particle is called a "hole" and it contributes toward the conduction of current. Holes can be formed by liberating electrons from covalent bonds, or by substituting semiconductor atoms with elements that have fewer valence electrons. In the latter case, vacancies are formed due to insufficient number electrons available for bonding. For example, holes are formed in Ge crystals (group IV) by substituting its atoms with elements from group III of the periodic table (e. g., Boron).

The energy states that an electron or hole can occupy in a semiconductor are specified in terms of the electronic band structure (figure 2.1). As shown, electrons increase in energy positively while a hole, negatively. The approximations used in generating the E-k diagram are presented next.

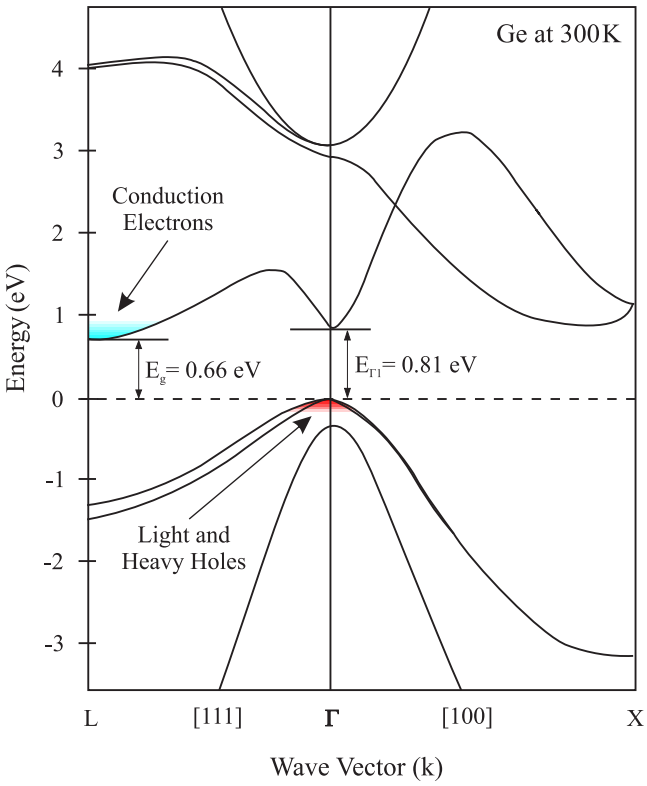


Figure 2.1: The energy band structure of Ge at 300 K calculated using nonlocal pseudopotential method[4]. Momentum is zero at Γ ($k = 0$), but increases on both directions along the abscissa, i. e., $\boldsymbol{p} > 0$ at 'L' and 'X'.

2.1.1 Germanium's Electronic Hamiltonian

The task of solving the wave equations for all the electrons in Ge ($\approx 10^{23}\,\mathrm{atoms/cm^3}$) involves several approximations that concern the microscopic behaviour of its carriers. In addition, the translational and rotational symmetry of the germanium crystal structure must be utilized in order to simplify computation. The latter is achieved using "Group theory", where most of its notations are found in the E-k diagram.

The first approximation used is the separation of electrons into two groups: the valence electrons and the core electrons[20]. The core electrons are those that reside in filled orbitals and are strongly bound to the nuclei. Therefore, only the valence electrons that participate in the formation of covalent bonds are considered. (The band structure in figure 2.1 accordingly shows only the allowed energy states for valence electrons.) The electronic Hamiltonian for the valence electrons of germanium can therefore be expressed as

$$\mathscr{H}_\mathrm{e} = \sum_i \frac{p_i^2}{2m_i} + \frac{1}{2}{\sum_{i,i'}}' \frac{1}{4\pi\varepsilon_0}\frac{q^2}{(\boldsymbol{r}_i - \boldsymbol{r}_{i'})^2} - \sum_{i,j}\frac{1}{4\pi\varepsilon_0}\frac{Z_j q^2}{(\boldsymbol{r}_i - \boldsymbol{R}_{j0})^2} \tag{2.1}$$

where $\boldsymbol{r}_i$ and $\boldsymbol{R}_{j0}$ denote the location of the i^{th} electron and j^{th} nucleus respectively, $\boldsymbol{p}_i$ the momentum operators of electrons, Z_j the atomic number of nucleus and q being the electronic charge. $\sum'$ only involves the summation over dissimilar pairs of indices. The first term denotes the total kinetic energy of the electrons while the remaining terms give the total potential energy (see equation 2.2).

The second approximation applied is the *mean-field approximation* where each electron is assumed to experience the same average, periodic potential $V(\boldsymbol{r})$ in the crystal lattice. With that, the motion of each electron in Ge can be described using the same Schrödinger equations written as

$$\mathscr{H}_{1e}\,\Phi_n(\boldsymbol{r}) = \left(\frac{p^2}{2m} + V(\boldsymbol{r})\right)\Phi_n(\boldsymbol{r}) = E_n\,\Phi_n(\boldsymbol{r}). \tag{2.2}$$

$\mathscr{H}_{1e}$, $\Phi_n(\boldsymbol{r})$ and E_n are respectively the one-electron Hamiltonian, the electron wavefunction and the energy of electron in eigenstates identified by the subscript n. (Each eigenstate can only accommodate up to two electrons of opposite spin as dictated by Pauli's exclusion principle.) Equation 2.2 can be solved using methods such as the pseudopotential method[4], the $\boldsymbol{k}\cdot\boldsymbol{p}$ method[8, 10] and the LCAO[20] method. We have adopted the $\boldsymbol{k}\cdot\boldsymbol{p}$ method in solving the equation 2.2 in Chapter 7 to estimate the energy separation between valence sub-bands (figure 7.14). The basis of using translational symmetry to simplify the task of solving equation 2.2 are elaborated next.

2.1.2 Translational Symmetry and Direct Lattice Sites

Certain translations along the crystal lattice of germanium will terminate at sites identical to the one prior translation (i. e., crystal structure remains invariant under translation). These translations occur along vectors that are the sum of integer multiples of the *primitive lattice vectors* with basis $\boldsymbol{a}_1$, $\boldsymbol{a}_2$ and $\boldsymbol{a}_3$. The direction for each basis is shown in figure 2.2 with vectors $\boldsymbol{a}_1 = (0, a/2, a/2)$, $\boldsymbol{a}_2 = (a/2, 0, a/2)$ and $\boldsymbol{a}_3 = (a/2, a/2, 0)$.

Therefore, locations invariant under translations in the crystal (direct lattice sites) can be defined by the set

$$\boldsymbol{R} = m\boldsymbol{a}_1 + n\boldsymbol{a}_2 + p\boldsymbol{a}_3 \tag{2.3}$$

where m, n and p are integers. $\boldsymbol{R}$ is important in describing translational symmetry operations in Ge, and is one of the parameter used in the *Bloch function*.

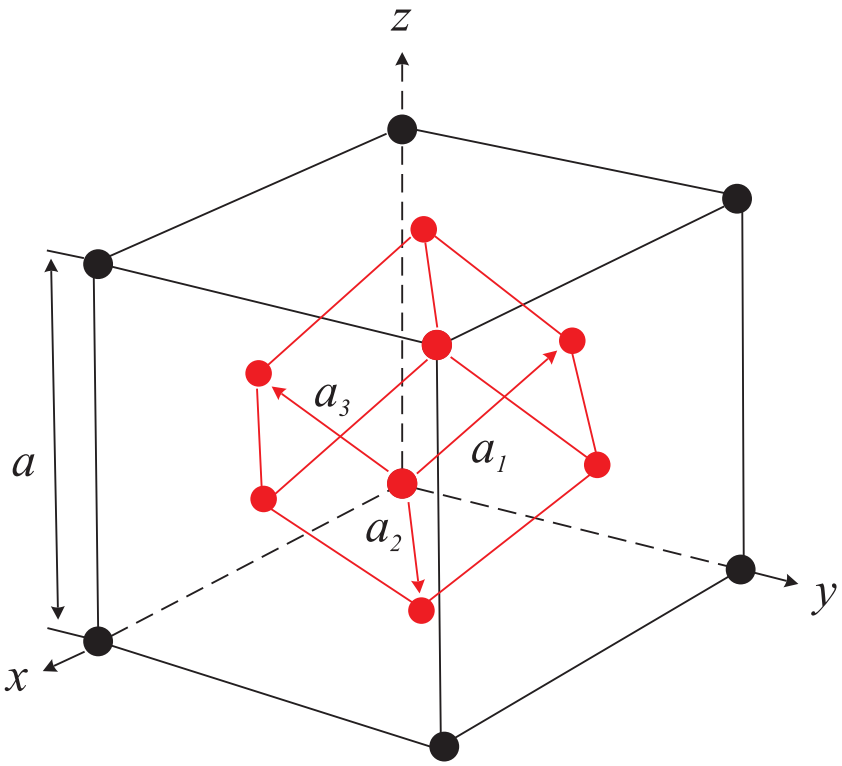


Figure 2.2: The Ge crystal structure enclosed by an imaginary cube with volume a^3 ($a = 5.64\,\text{Å}$). This cube is known as the face-centred cube (fcc), and it contains the smallest assembly of atoms required to form a complete crystal. Shown also is the set of primitive lattice vectors a_1, a_2 and a_3 for Ge (all elements in the fcc consist only of Ge atoms).

2.1.3 The Bloch Function

The Bloch function can be used to describe the wavefunction of particles under the influence of periodic potentials. In one-dimension, it has the form

$$\psi_k(x) = \exp(\mathrm{i}kx)u_k(x) \tag{2.4}$$

To see how the Bloch function can be used in describing the wavefunctions of electrons in Ge, we first invoke the mean-field approximation (section 2.1.1), and, for simplicity, limit equation 2.2 to one-dimension (easily generalized to three). All of the electrons now experience the same periodic potential $V(\boldsymbol{x})$, and this periodicity is shared by the function $u_k(x)$, i. e., $u_k(x) = u_k(x + n\boldsymbol{R})$ where $\boldsymbol{R}$ is the vector that gives invariant translation through the lattice. ($u_k(x)$ is lattice periodic. In one dimension, it is expressed as a general solution for the second order, linear, homogeneous differential equation, with boundary conditions set by $V(\boldsymbol{x})$. Thus it has the shape of a damped harmonic oscillator [17]. In solving equation 2.2 using the $\boldsymbol{k} \cdot \boldsymbol{p}$ method, only

data from cyclotron resonance experiments in the form of momentum matrix elements, together with data from absorption and reflectivity measurements are needed [8].) By multiplying equation 2.4 with $\exp(-\mathrm{i}\omega t)$, we have an equation that represents a plane wave with amplitude modulated by $u_k(x)$. Now if x changes to $x + \boldsymbol{R}$, periodicity causes $\psi_k(x)$ in equation 2.4 to vary according to

$$\psi_k(x + \boldsymbol{R}) = \exp(\mathrm{i}k\boldsymbol{R})\psi_k(x). \tag{2.5}$$

The implications of equation 2.5 are conveyed by defining a translation operator $T_{\boldsymbol{R}}$ such that

$$T_{\boldsymbol{R}}\psi_k(x) = \psi_k(x + \boldsymbol{R}) = \exp(\mathrm{i}k\boldsymbol{R})\psi_k(x). \tag{2.6}$$

From equation 2.6, we can see that $\psi_k(x)$ acts as an eigenfunction of $T_{\boldsymbol{R}}$ with eigenvalue $\exp(\mathrm{i}k\boldsymbol{R})$. Since $\mathscr{H}_{1e}$ in equation 2.2 is invariant under the translation set $\boldsymbol{R}$, $\mathscr{H}_{1e}$ also commutes with $T_{\boldsymbol{R}}$. It follows then that the eigenfunction of $T_{\boldsymbol{R}}$ can also be used to express the eigenfunctions of $\mathscr{H}_{1e}$. As a result, we can define $\Phi(\boldsymbol{x})$ in equation 2.2 (assuming one-dimension) as the sum of the Bloch functions:

$$\Phi(x) = \sum_k A_k\psi_k(x) = \sum_k A_k \exp(\mathrm{i}kx)u_k(x). \tag{2.7}$$

k is the index representing the wave vectors of plane waves and A_k are constants. (Not all values of k are unique as values identical to $k + 2\mathrm{n}\pi/\boldsymbol{R}$ also satisfies equation 2.4.) k can also be replaced by $k' = k - 2\mathrm{n}\pi/\boldsymbol{R}$, with integer 'n' chosen to limit k' within the range $[-\pi\boldsymbol{R}, \pi\boldsymbol{R}]$. The k-space occupied by k' is known as the *first Brillouin Zone*.

2.1.4 The First Brillouin Zone and the Reduced Zone Scheme

When translational symmetry is utilized in solving equation 2.2 (section 2.1.3), $\psi_k(x)$ by definition must be equal to $\psi_k(x + \boldsymbol{R})$. This is achieved if $k \cdot \boldsymbol{R}$ is a multiple of 2π. As a consequence, simplification is only possible if k is computed along vectors with period $k \cdot \boldsymbol{R} = 2\pi$. These vectors can be obtained from the primitive lattice via the following operation:

$$\boldsymbol{b}_i = 2\pi\frac{(\boldsymbol{a}_j \times \boldsymbol{a}_k)}{(\boldsymbol{a}_1 \times \boldsymbol{a}_2) \cdot \boldsymbol{a}_3} \tag{2.8}$$

where i,j and k are respectively the cyclic permutation indices 1, 2 and 3 of direct basis in set $\boldsymbol{R}$. As a result, three new basis vectors called the *primitive reciprocal lattice vectors* are produced. These reciprocal basis are denoted $\boldsymbol{b}_1$, $\boldsymbol{b}_2$ and $\boldsymbol{b}_3$, corresponding to, respectively, the direct basis vectors with similar index in set $\boldsymbol{R}$. From equation 2.8,

the vectors for each basis in the reciprocal lattice are accordingly $\boldsymbol{b}_1 = (\frac{2\pi}{a})(-1,1,1)$, $\boldsymbol{b}_2 = (\frac{2\pi}{a})(1,-1,1)$ and $\boldsymbol{b}_3 = (\frac{2\pi}{a})(1,1,-1)$. In addition, they have properties such that $\boldsymbol{a}_i \cdot \boldsymbol{b}_i = 2\pi$, $\boldsymbol{a}_i \cdot \boldsymbol{b}_j = 0$, and so on. The primitive reciprocal lattice vectors thus form a new vector set expressed as

$$\boldsymbol{G} = h\boldsymbol{b}_1 + k\boldsymbol{b}_2 + l\boldsymbol{b}_3 \tag{2.9}$$

where h, k and l are constants. Since $\boldsymbol{G} \cdot \boldsymbol{R} = 2\pi \times$ integer, one could then compute k along vector $\boldsymbol{G}$ (equation 2.5, 2.6 and 2.7).

As mentioned previously, the first Brillouin Zone corresponds to the k-space in which k lies within the interval $[-\pi/\boldsymbol{R}, \pi/\boldsymbol{R}]$. If k also conforms to $\boldsymbol{G}$ vectorially, the restricted k-space for Ge in three dimensions, enclosed inside a fcc cube is shown in figure 2.3[19]. The space occupied is known as the Wigner-Seitz (WS) cell or the first Brillouin Zone.

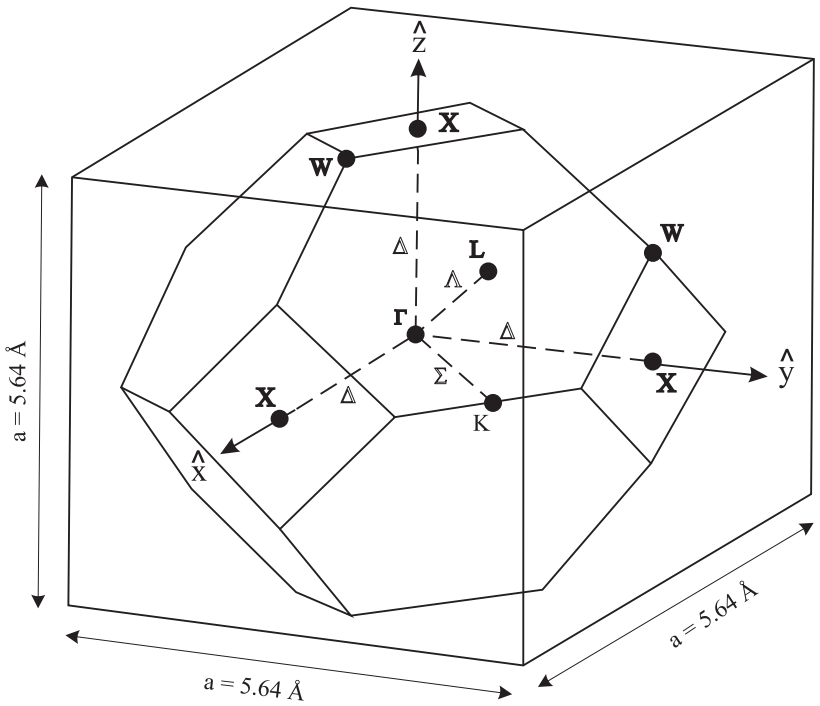


Figure 2.3: Germanium's primitive cell encasing a truncated octahedron called the Wigner-Seitz (WS) cell. The volume occupied by the WS cell is known as the first Brillouin Zone.

High symmetry points and vectors inside the WS cell are labelled using Greek letters while Roman letters are used for symmetry points on the surfaces of the cell. The crystallographic points and axes used in the Brillouin Zone are compared with conventions used by Miller Indices in table 2.1.

Point	Degeneracy	Axis
Γ, $(0,0,0)$	1	
X, $2\pi/a(\pm1,0,0)$, $2\pi/a(0,\pm1,0)$, $2\pi/a(0,0,\pm1)$	6	$\Delta, \langle1,0,0\rangle$
L, $2\pi/a(\pm1/2,\pm1/2,\pm1/2)$	8	Λ, $\langle1,1,1\rangle$
K, $2\pi/a(\pm3/4,\pm3/4,0)$, $(0,\pm3/4,\pm3/4)$, $(\pm3/4,0,\pm3/4)$	12	Σ, $\langle1,1,0\rangle$

Table 2.1: The axes and edges of Brillouin zone of diamond lattice

The utilization of translational symmetry and the Bloch function in solving equation 2.2 allows the mapping of free carrier (electrons and holes) energy states within the WS cell. The result of this approach is shown in figure 2.1, with energy levels calculated as a function of k within the first Brillouin Zone. Limiting k in such a way is known as the *reduced zone scheme*, and has the advantage of producing electronic band structures that are more compact[20].

2.2 Carrier Transitions in Intrinsic Germanium

The E-k diagram will now be used for illustrating carrier transitions within Ge. Discussions in this section are confined to perfect crystalline intrinsic Ge semiconductors with no defects or impurities present.

In Ge, electrons can be liberated from covalent bonds via methods such as optical illumination, thermal agitation and—with pn junction structures fabricated—even electrical injection[20]. For each electron liberated from the valence band (VB), a hole is formed in its place. The minimum energy required by electrons to be excited to the conduction band (CB) in Ge is 0.66 eV[17], which corresponds to the diffference in energy between the lowest point in the conduction band and the highest peak of the valence band located at Γ (figure 2.1). This energy difference is known as the semiconductor bandgap (E_g). Within this gap, no energy state can exist unless defects or impurities are present in the crystal.

The total energy an electron can have in the conduction band is[1]

$$E = E_\mathrm{c} + \frac{|\boldsymbol{p}|^2}{2m^*_\mathrm{ce}} \tag{2.10}$$

where E_c is the minimum energy at the CB that varies with wavevector k (directionally dependent as a result of the crystalline nature of Ge crystals); $\boldsymbol{p}$ is the momentum of electrons; and m^*_ce is the electron conductivity effective mass defined in equation 2.15. Electrons that reside in the conduction band has two properties. First is the conserva-

tion of momentum. For each point beyond the centre of the Wigner -Seitz cell, non-zero momentum states exist for each corresponding energy levels in the CB[1]. Electrons therefore require both the right energy and momentum to be excited to the CB. *Indirect transition* occurs when momentum must be conserved with the help of phonons for transition to take place. However, if electrons are excited with energies greater than 0.81 eV ($\approx 1.55\,\mu m$), direct transitions between the VB and the CB can take place at $k = \Gamma$ without additional change in the momentum of electrons[18]. This transition explains the the steep rise in absorption coefficient around $1.55\,\mu m$ in figure 2.6.

Momentum is conserved via the emission or absorption of phonons—a quantum of lattice vibration with momentum the order of h/a (a being the lattice constant)[14]. Their large momentum compared to that of photons (h/λ), or any other scatterers such as electrons and holes make them the dominant mediator for transfering momentum to electrons involved in indirect transitions[14]. Six modes of phonons exist in Ge: 3 optical phonons and 3 acoustic phonons. Both optical and acoustic phonons have modes traversing longitudinally along the $< 1, 1, 1 >$ direction of the crystal lattice (LO and LA), and two phonons in transverse mode (TA and TO). Two longitudinal acoustic phonons located at the $<1, 1, 1>$ edge of the WS cell are involved in indirect transitions in Ge. These phonons have energies comparable to temperatures 0.0075 eV and 0.0275 eV[12].

This brings us to the second implication of equation 2.10: So long as Pauli's exclusion principle is not violated, the total energy of an electron excited to the conduction band is independent of the direction of the momentum vector. Only the magnitude of the momentum vector is required by electrons to participate in indirect transitions. (The momentum in equation 2.10 refers to the motion of electrons through periodic lattice potentials. Since the optical and electronic properties of crystals in general exhibit anisotropy with respect to different crystallographic directions, it would be normal to assume that certain components of the momentum vector (i. e., specific direction of motion) must exist to enable electrons to occupy discrete energy levels in the conduction band. This is not the case.) This is the reason for the large number of momentum states with similar energy in the conduction and valence band.

The number of states that electrons can occupy in the CB without causing degeneracy is given by the CB effective density of states formula:

$$N_{\mathrm{c}} = 2\left(\frac{2\pi m^*_{\mathrm{de}} kT}{h^2}\right)^{3/2} M_{\mathrm{C}}. \tag{2.11}$$

M_{C} is the number of equivalent minima in the CB and m^*_{de} the electron density of state effective mass given as $(m_{\mathrm{l}}^* m_{\mathrm{t}}^{*2})^{1/3}$. In the same way, the VB effective density of states for germanium (N_v) is given as:

$$N_{\mathrm{v}} = 2\left(\frac{2\pi m^*_{dh} kT}{h^2}\right)^{3/2}, \tag{2.12}$$

while the energy a hole can have in the VB is

$$E = E_{\mathrm{v}} - \frac{|\boldsymbol{p}|^2}{2m^*_{\mathrm{h}}}. \tag{2.13}$$

m_{dh} is the hole density of state effective mass given as $(m_{\mathrm{lh}}^{*\,3/2} + m_{\mathrm{hh}}^{*\,3/2})^{2/3}$ (section 2.2.1). Holes, which represent the movement of groups of bound electrons surrounding mobile, broken bonds in the crystal lattice, gain energy negatively (by convention) through interactions with phonons.

At 300 K, N_{c} and N_{v} approaches $10^{19}\,\mathrm{cm}^{-3}$ and $4 \times 10^{18}\,\mathrm{cm}^{-3}$ respectively. Such high density of states in each band permit the generation of a large number of free carriers that induces IR absorption in Ge between $2 - 14\,\mu$m (section 2.2.1 and section 2.2.2).

2.2.1 Free-Carrier Absorption

Photons with energy less than E_{g} can excite free carriers to higher energy levels within the same band. This process occurs both in the conduction and valence band, giving rise to absorption known as the free-carrier absorption (FCA)[18]. As with indirect transitions, carrier excitation in FCA must be assisted by phonons in order to conserve momentum (figure 2.4).

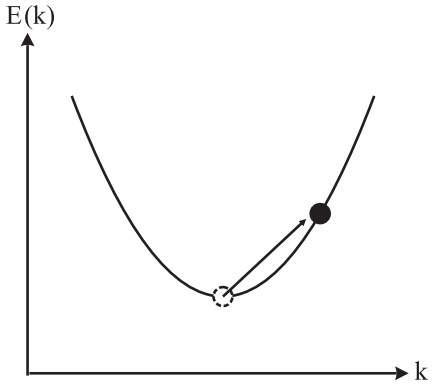


Figure 2.4: Transition of a free electron to a higher energy state within the same valley in the conduction band. Transitions with larger Δk requires interaction with phonons with larger momentum

FCA can be treated using the quantum theory of transition probabilities between states of different energies[7]. However, a semi-classical approach giving results similar to quantum theory is usually adopted because of its reduced complexity. The latter is known as the Drude-Zener theory of FCA, and it assumes the presence of spherically symmetric, nondegenerate energy surfaces. For Ge, this region is located along the (111) direction of **k** space, in which the energy surfaces are prolate spheroidal for the conduction band[11]. The Drude-Zener FC absorption coefficent (α_{fc}) for semiconductors can be expressed as[18]:

$$\alpha_{\mathrm{fc}} = \frac{q^3 \lambda^c}{4\pi^2 \mathrm{c}^3 \varepsilon_0 n_{\mathrm{index}}} \left[\frac{n}{m_{\mathrm{n}}^{*2}} + \frac{p}{m_{\mathrm{p}}^{*2}} \right] \tag{2.14}$$

where n_{index} is the refractive index, n and p the electron and hole density respectively, and m^* the effective mass of carriers (subscript 'n' for electrons and 'p' for holes). The electron mass used is the electron conductivity effective mass written as[11]:

$$\frac{1}{\langle m^* \rangle_{ce}} = \frac{1}{3} \left(\frac{1}{m_{\mathrm{l}}^*} + \frac{2}{m_{\mathrm{t}}^*} \right) \tag{2.15}$$

m_{ce} is an average between the longitudinal electron effective mass (m_{l}) travelling in direction $<1,1,1>$, and the transverse electron effective mass (m_{t}). In the absence of a magnetic field, the heavy hole mass, m_{hh}^*(section 2.2.2), is used for m_{p} in evaluating the FC absorption coefficient[6].

Semiconductor FCA spectra, when not overshadowed by stronger absorption mechanisms, are often structureless (n-type Ge in figure 2.5), and exhibit absorption dependency that is proportional to two parameters: the wavelength of incident radiation λ^c, and the concentration of free carriers (equation 2.14). At shorter wavelengths, provided that photon energies are less than E_{g}, interaction with acoustic and optical phonons are more probable, causing α_{fc} to vary according to $\lambda^{1.5}$ and $\lambda^{2.5}$ respectively[3]. At longer wavelengths, FCA dependence on phonons are reduced as smaller Δk steps (smaller momentum difference) are made by excited carriers (figure 2.4). Under this condition, a λ^2 increase for α_{fc} is generally observed for most semiconductors[18].

To illustrate the wavelength dependency of α_{fc} between $2 - 14\,\mu$m, the IR spectra of a hand-polished n-type Ge wafer with dopant concentration $\approx 2 \times 10^{18}\,\mathrm{cm}^{-3}$ was measured using a Fourier transform IR spectrometer (Spectrum GX FTIR). α_{fc} was then evaluated using the equation[2]:

$$\alpha_{\mathrm{fc}} = -\frac{1}{\mathrm{d}} \left(\frac{\sqrt{(1-R)^4 + 4T^2R^2} - (1-R)^2}{2TR^2} \right). \tag{2.16}$$

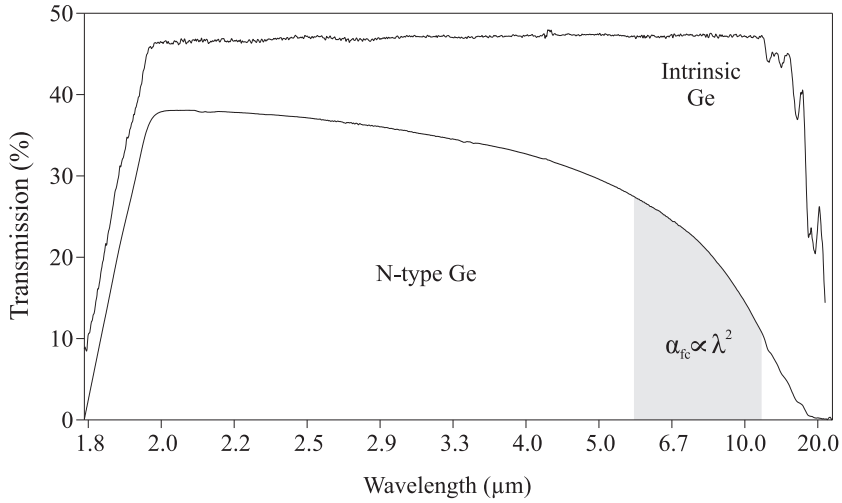


Figure 2.5: Transmission spectra of a 3 mm thick intrinsic Germanium and an n-type germanium with dopant concentration $\approx 5.3 \times 10^{16}\,\mathrm{cm}^{-3}$. At wavelengths above 6 μm, α_{fc} starts to behave as λ^2.

R, T and d are respectively the reflectance, transmittance and the thickness of the sample. R was approximated to $\frac{(n_{\mathrm{air}}-n_{\mathrm{Ge}})^2}{(n_{\mathrm{air}}+n_{\mathrm{Ge}})^2}$ since the extinction coefficients for Ge between 2 to 14 μm are negligable[15]. The refractive index used in R was calculated using a Sellmeier-type equation[15]

$$n_{\mathrm{Ge}}{}^2 = A + B\lambda^2/(\lambda^2 - C) + D\lambda^2/(\lambda^2 - E) \tag{2.17}$$

where A, B, C, D and E are the temperature-dependent coefficients (T is the temperature here):

$$\begin{aligned} A &= -6.040 \times 10^{-3}\,T + 11.05128 \\ B &= 9.295 \times 10^{-3}\,T + 4.00536 \\ C &= -5.392 \times 10^{-4}\,T + 0.5999034 \\ D &= 4.151 \times 10^{-4}\,T + 0.09145 \\ E &= 1.51408T + 3426.5 \end{aligned} \tag{2.18}$$

For an absorbing medium, the transmittance T is given as[16]:

$$T = \frac{(1-R)^2 e^{-\alpha d}}{1 + R^2 e^{-2\alpha d} + 2Re^{-\alpha d}\cos(\phi)}, \tag{2.19}$$

and was obtained directly from the spectrometer with resolution set to $16\,\mathrm{cm}^{-1}$. ϕ is the ratio between f/f_1 where $f = 2\pi/\lambda$ and $f_1 = 1/(2n_{\mathrm{Ge}}d)$. Since the resolution set is $> 1/(2n_{\mathrm{Ge}}d)$ for a sample 340 μm thick, the fine-structured oscillations in T, represented by the term $\cos(\phi)$, are not resolved[16].

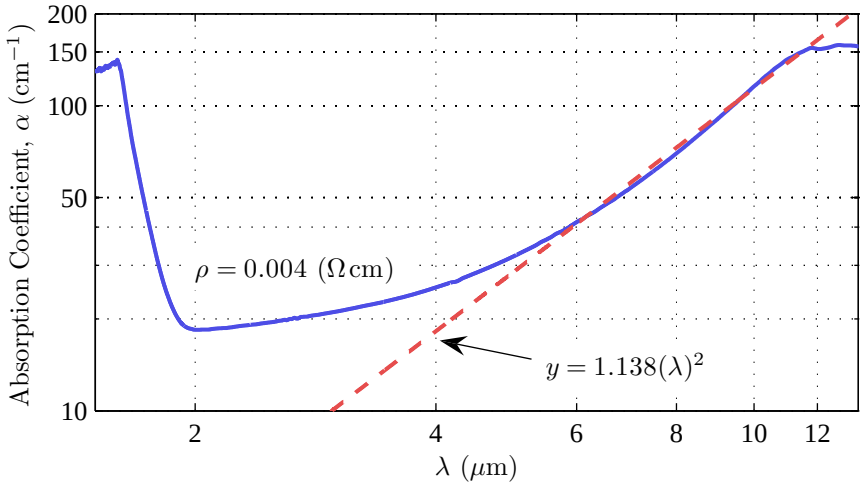


Figure 2.6: The absorption coefficient of an n-type Ge wafer with dopant concentration $\approx 2 \times 10^{18}\,\mathrm{cm}^{-3}$. The sharp rise about $1.55\,\mu$m is due to direct electronic transitions from the VB to the CB. At wavelength longer than $6.0\,\mu$m, α_{fc}varies with λ^2

Ionised impurities also contribute to the scattering of carriers, causing α_{fc} to vary between λ^3 and $\lambda^{3.5}$. This discrepancy can be attributed to the difference in the chemical nature of impurities[14]. As an example, the absorption coefficients observed for Ge samples separately doped with As, P and Sb, are found to behave as:

$$\alpha_{\mathrm{As}} > \alpha_{\mathrm{P}} > \alpha_{\mathrm{Sb}} \tag{2.20}$$

In general, all three forms of scattering mechanisms are present at any one time, thus α_{fc} can be written as a weighted sum of three processes[14]:

$$\alpha_{\mathrm{fc}} = \mathrm{A}\lambda^{1.5} + \mathrm{B}\lambda^{2.5} + \mathrm{C}\lambda^{d}, \quad 3 \leq d \leq 3.5 \tag{2.21}$$

where A, B and C are constants. As a result, α_{fc} variation with λ cannot be predicted accurately using simple power law when a broad spectrum is considered. However, FCA does increase rapidly with wavelength.

For germanium, a much stronger absorption mechanism overlaps FCA at the modulator's operating band ($8 - 14\,\mu$m). This will be explored in Section. 2.2.2. Despite that, the contribution of FCA towards the overall absorption between 8 and $14\,\mu$m starts to become significant; therefore, its effect must be included in the evalutation of modulator performance(section 2.3).

2.2.2 Intervalence Carrier Transitions

Germanium's valence band consist of several subbands[19]—three of which are responsible for the characteristic transmission peaks observed in p-type Ge between $2-14\,\mu$m (figure 2.8)[11]. These subbands are shown in figure 2.7. Close to Γ, the top two subbands can be approximated by parabolic bands with different curvatures[19]. They are named the heavy hole band (HH) and the light hole band (LH) repectively, attributed to the difference in the energy gradient between the two with $(\frac{\partial^2 E}{\partial k^2})_{\mathrm{LH}} > (\frac{\partial^2 E}{\partial k^2})_{\mathrm{HH}}$. The third subband splits from the rest due to spin orbit interaction and is known as the spin-orbit split-off band (SO)[18]. These subbands give rise to different hole effective mass that are in general tensorial and can be defined as[19]

$$\frac{1}{m_{\mathrm{ij}}} = \frac{1}{\hbar}\frac{\partial^2 E(k)}{\partial k_{\mathrm{i}} \partial k_{\mathrm{j}}}. \tag{2.22}$$

For our purposes, only the heavy hole mass ($m^*_{\mathrm{hh}} \approx 0.28\,m_0$) and the light hole mass ($m^*_{\mathrm{lh}} \approx 0.04\,m_0$) are required for evaluating the contribution of holes toward the optical and electrical properties of Ge[19]. Values for m^*_{hh} and m^*_{lh} were derived from cyclotron resonance experiments, which provide the most direct method in determining carrier effective mass[18].

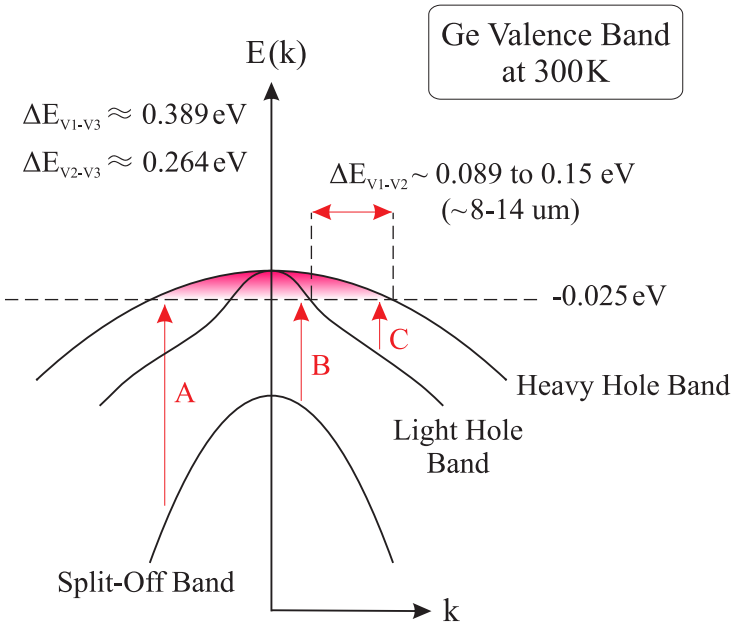


Figure 2.7: Germanium's valence band structure near Γ at 300 K (not to scale). Interband transitions A and B give rise to IR absorptions centred at $3.18\,\mu$m and $4.68\,\mu$m respectively, while transition C results in the broad absorption between $8-14\,\mu$m

When holes populate the valence band, three types of electronic transitions are

possible between the subbands: (a) the SO band to the HH band; (b) the SO band the LH band; and (c) the LH band to the HH band (figure 2.7)[14]. The first two transitions give rise to IR absorptions centred at 3.18 μm and 4.68 μm respectively, while the third results in the broad absorption between $8 - 14\,\mu$m (figure 2.8). The magnitude of these absorptions are proportional to the concentration of holes in Ge. As a result, they can be significantly reduced by introducing donor atoms at densities above thermally generated free carrier concentrations—thereby turning the semiconductor n-type[18]. For non-degenerate n-type Ge crystals, donor concentrations obey the formula[19]

$$n_{\mathrm{n}} = \frac{n_{\mathrm{i}}^2}{p_{\mathrm{n}}} \tag{2.23}$$

where n_{n} is the density of electrons per unit volume (cm^{-3}) predominantly contributed by donor atoms, p_{n} the hole density, and n_{i} is the intrinsic carrier concentration. Thus the increase in n_{n} reduces p_{n}, which in turn lowers the magnitude of absorption peaks related to intervalence transitions. The IR spectra of an intrinsic Ge ($<10^{10}\,\mathrm{cm}^{-3}$ impurities) and a p-type Ge are compared in figure 2.8.

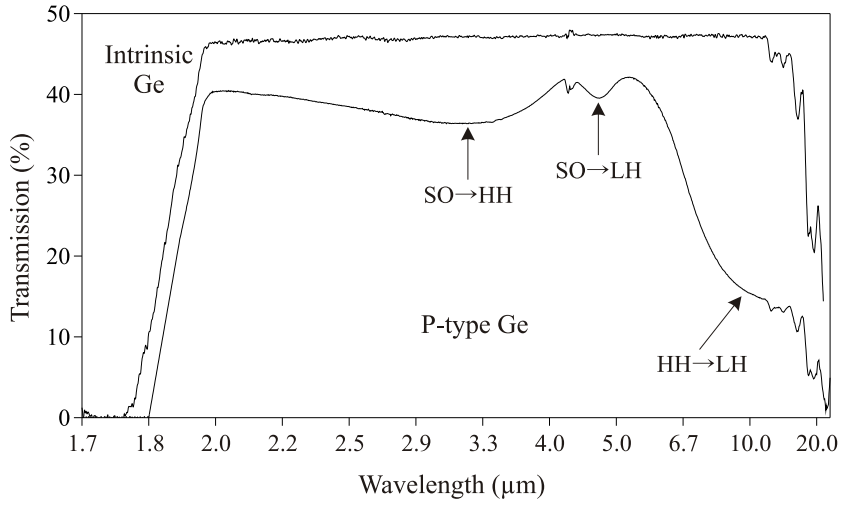


Figure 2.8: Transmission spectra for a p-type Ge sample. Beyond 14 μm, multiphonon absorptions start to become dominant[14]

Because intervalence transitions are direct[14], they are more probable compared to FCA carrier transitions (section 2.2.1). Intervalence transitions therefore are the main contributor toward the broad absorption observed in p-type Ge between $8 - 14\,\mu$m.

The relaxation time for electrons that are excited to the LH band from higher energy bands (e. g., HH band) is in the order of 10^{-12} seconds. Therefore, absorption due to intervalence transitions A, B and C only saturates under intense conditions such as the absorption of MW power levels[6].

Changes in the temperature affect the position of absorption peaks shown in figure 2.8. At low temperatures, the fermi level moves away from the valence band. This causes transition A and C to shift to lower energies while transition B to higher energies[14].

2.3 Carrier Injected IR Modulation

By fabricating Ge PiN diodes, electrons and holes of similar concentrations can be injected into the intrinsic 'i' region by applying a forward voltage on the diode[19]. Between $8 - 14\,\mu$m, the absorption coefficient in this region can be expressed as[13]:

$$\alpha = \alpha_{\mathrm{n}} + \alpha_{\mathrm{p}} + \alpha_{\mathrm{HH}\rightarrow\mathrm{LH}}. \tag{2.24}$$

α_{n} and α_{p} are contributions from the electron and hole FCA coefficients (section 2.2.1), while $\alpha_{\mathrm{HH}\rightarrow\mathrm{LH}}$ is the absorption associated with intervalence transition between the HH band and the LH band.

Carrier injected IR modulation using Ge PiN diodes has been demonstrated by us, and some of the results are presented in figure 2.9. Details of the fabrication of the diodes and the temporal analysis of their performance at $10.59\,\mu$m are given in chapters 6 and 7 respectively.

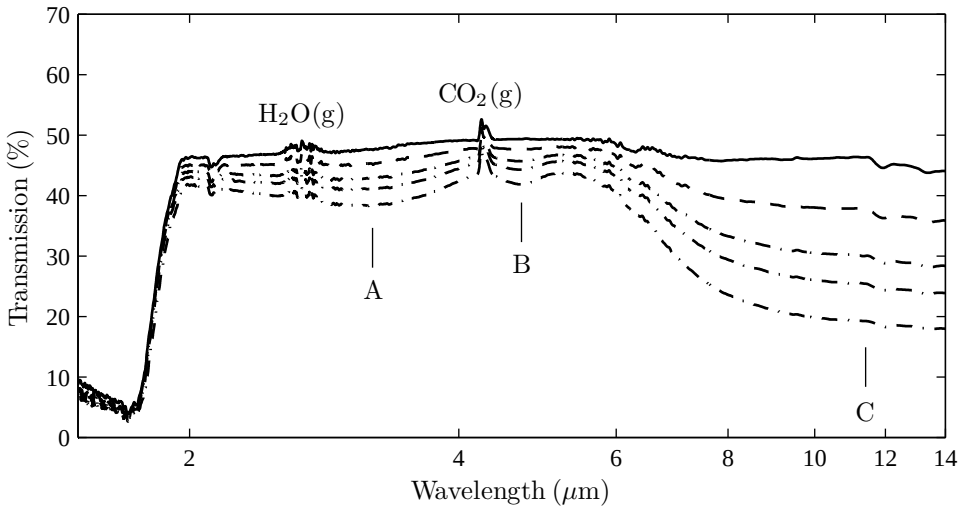


Figure 2.9: The transmission spectra through Ge PiN diode MI2-50-3 (table 7.1) operated at currents 0, 0.1, 0.4, 0.8, 1.2 and 2.0 A (the depth of modulation increases with current). Absorption peaks A, B and C are attributed to intervalence transition describe in section 2.2.2

To estimate the number of carriers involved in IR modulation, the overall absorption cross section for electrons $\sigma_{\mathrm{e}}\,(\mathrm{cm}^2)$ and holes $\sigma_{\mathrm{h}}\,(\mathrm{cm}^2)$ are used. This approach takes

into account absorptions due to FCA and intervalence transitions simultaneously. In addition, our work was focussed on the CO_2 laser operating wavelength at $10.59\,\mu$m since σ_e and σ_h at that wavelength have been measured[9].

The transmission through intrinsic Ge as a function of carrier area density at $10.59\,\mu$m can then be expressed as

$$T = T_0 \exp[-(\sigma_e + \sigma_h) \cdot A_{h,e}]. \tag{2.25}$$

$A_{h,e}$ is the integral of either type of injected carrier density through the entire depth of the diode in the $\hat{z}$ direction (figure 1.1), i. e., $A_{h,e} = \int_0^d N(x)\,dx$, where $N(x)$ is either the density of electrons or holes per cm^3. (Due to charge neutrality requirement in intrinsic Ge crystals, the density of electrons and holes injected must be equal[19]. Hence equation 2.25 only requires the area density of one type of carrier.)

At $10.59\,\mu$m, the absorption cross section for electrons and holes in Ge are 0.34×10^{-16} and $5.33 \times 10^{-16}\,cm^{-2}$ respectively. This gives a ratio of 15.7 for σ_h/σ_e. Furthermore, the trend where σ_h is larger than σ_e is observed for all absorption regions associated with intervalence transitions in Ge[18].

Therefore, the presence of holes are crucial for the performance of our IR modulator because the bulk of its absorptions between $8 - 14\,\mu$m are mainly due to intervalence carrier transitions. Methods employed in optimizing the electrical and optical performance of the diode modulator are presented in chapter 5.

2.4 Summary

The bulk of IR absorption observed in non-degenerate Ge between $2 - 14\,\mu$m is caused by two different electronic transitions: the intraband free carrier transitions (FCA), and the intervalence carrier transitions. The latter dominates in regions centred around 3.18, 4.68 and $10\,\mu$m, where its contribution is several times larger compared to FCA.

Carrier transition induced IR absorption has been utilized to produce solid-state modulators in the form of PiN diodes. Absorption was successfully demonstrated by electrically injecting carriers into Ge substrate.

Bibliography

[1] R. B. Adler, A. C. Smith, and R. L. Longini. *Introduction to Semiconductor Physics*, volume 1. John Wiley & Sons, New York, 1964.

[2] N. P. Barnes and M. S. Piltch. Temperature-dependent sellmeier coefficients and nonlinear optics average power limit for germanium. *Journal of the Optical Society of America*, 69(1):178–180, 1979.

[3] M. Becker and H. Y. Fan. Proceedings of the reading conference. In H. K. Henisch, editor, *Semiconducting Materials*, London, 1951. Butterworth Scientific Publications.

[4] J. R. Chelikowsky and M. L. Cohen. Nonlocal pseudopotential calculations for the electronic structure of eleven diamond and zinc-blende semiconductors. *Physical Review B*, 14(2):556–582, 1976.

[5] G. Dresselhaus, A. F. Kip, and C. Kittel. Cyclotron resonance of electrons and holes in silicon and germanium crystals. *Physical Review*, 98(2):368–384, 1955.

[6] P. D. Fairley. *Novel Solid State Modulator for the Infrared: The Germanium Chopper*. PhD thesis, University of Southampton, 2000.

[7] H. Y. Fan. Infra-red absorption in semiconductors. *Reports on Progress in Physics*, 19:107–155, 1956.

[8] W. Fawcett. Valence band structure of germanium. *Proceedings of the Physical Society*, 85:931–944, 1965.

[9] C. J. Hutchinson, C. Lewis, J. A. Savage, and A. Pitt. Surface and bulk absorption in germanium at 10.6 μm. 1982.

[10] E. O. Kane. Energy band structure in p-type germanium and silicon. *Journal of Physics and Chemistry of Solids*, 1:82–99, 1956.

[11] A. H. Khan. Theory of the infrared absorption of carriers in germanium and silicon. *Physical Review*, 97(6):1647–1652, 1955.

[12] G. G. MacFarlane, T. P. McLean, J. E. Quarrington, and V. Roberts. Fine structure in the absorption-edge spectrum of ge. *Physical Review*, 108(6):1377–1383, 1957.

[13] I. M. Nesmelova, N. I. Astaf'ev, and E. A. Nesmelov. Resistivity dependence of the absorption coefficient of crystalline germanium in the ir region. *Optical Material Science and Technology*, 74(1):71–74, 2007.

[14] J. I. Pankove. *Optical Processes in Semiconductors*. Dover Publications, New Jersey, 2^{nd} edition, 1975.

[15] R. F. Potter. Semiconductors. In E. D. Palik, editor, *Handbook of Optical Constants of Solids II*. Academic Press, 1991.

[16] D. K. Schroder. *Semiconductor Material and Device Characterization*. John Wiley & Sons, New Jersey, 2nd edition, 2006.

[17] K. Seeger. *Semiconductor Physics: An Introduction*. Springer-Verlag, Berlin, 6^{th} edition, 1997.

[18] R. A. Smith. *Semiconductors*. Cambridge University Press, Cambridge, 2^{nd} edition, 1978.

[19] S. M. Sze and K. K. Ng. *Physics of Semiconductor Devices*. John Wiley & Sons, New Jersey, 3^{rd} edition, 2007.

[20] P. Y. Yu and M. Cardona. *Fundamentals of Semiconductor*. Springer, Berlin, 1^{st} edition, 1996.

Chapter 3

Electrical Properties of Germanium

Well established empirical models that describe the electrical properties of Ge are presented in this chapter. These models are implemented by our choice of device simulator, ATLAS, which was used to predict the electrical behaviour of our modulator in chapter 5 [35]. The use of empirical models is a practice universally adopted by many workers in the field of semiconductor physics, owing to the simplicity and accuracy offered [12, 39]. For our interest, only models pertinent to the functionality of the modulator are included. (Whenever present, syntax used in ATLAS are printed using "typewriter" family font, e. g., `NC300` and `NV300`.)

3.1 Semiconductor Bandgap (E_g)

The semiconductor bandgap E_g is smallest energy difference between the conduction band (CB) minima and the valence band (VB) maxima [36] (see figure 2.1); it is also one of the most frequently used parameter for evaluating the electrical and optical properties of semiconductors [1, 25].

E_g decreases with increasing temperature, doping concentration and even carrier concentration; with the last two terms evident only under degenerate conditions. In the case of high doping concentrations, shrinkage in the energy gap is due to interactions between ionized impurities and free carriers [2], while interactions between electrons and holes generated by external sources are responsible for the the narrowing of E_g at degenerate levels of carrier concentrations [16].

3.1.1 Dopant Dependent Energy Gap

The change in E_g due to impurities with shallow energies (i. e., energies close to the CB or VB) can be approximated using the following equation [37]:

$$\Delta E_g(N_{e,h}) = -\frac{3q^2}{16\pi\varepsilon_s}\sqrt{\frac{q^2 \cdot N_{e,h}}{\varepsilon_s kT}}. \tag{3.1}$$

where $N_{e,h}$ is the doping density for donors (subscript e) or acceptors (subscript h), q is the electronic charge, ε_s is the semiconductor permittivity, k is Boltzmann's constant and T is the temperature in Kelvin. For Ge (relative permittivity $\varepsilon_r = 16$) this expression further reduces to:

$$\Delta E_g(N_{e,h}) = -14.1\sqrt{\frac{N}{10^{18}\ \mathrm{cm}^{-3}}}\ \mathrm{meV}. \tag{3.2}$$

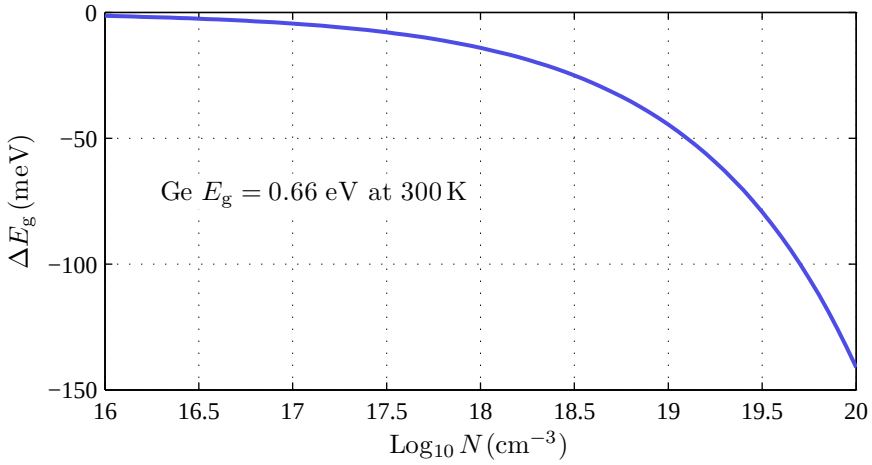


Figure 3.1: The change in the energy gap of germanium as a function of doping concentration (N)

As can be seen in figure 3.1, changes in E_g only start to become significant at dopant densities above $10^{18}\,\mathrm{cm}^{-3}$. Degeneracy due to high donor concentrations can be reasoned using Pauli's exclusion principle [1]. Recalling the conductivity effective mass (m^*_{ce}) defined by Eq. 2.15, and taking into account Heisenberg's uncertainty principle that relates quantitatively the lack of precision between the location, ($\Delta \boldsymbol{r}$), and the momentum of a carrier ($\Delta \boldsymbol{p}$) by the formula $\Delta \boldsymbol{r} \cdot \Delta \boldsymbol{p} \geq h$, a free electron in a semiconductor has an "effective size" of $\Delta \boldsymbol{r} \approx \frac{h}{\sqrt{3m^*_{ce}kT}}$ [1]. Quantum mechanically, this means that the *de Broglie* waves extends over the diameter of this size. If dense carriers are present in the crystal lattice, the overlapping of carrier waves will cause

significant quantum interactions to take place in order to prevent the violation of the exclusion principle. Degeneracy thus occurs[1], altering the behaviour of semiconductors to become more metal-like [1]. If dopants are uniformly distributed, one can estimate the minimum density of carriers (n_{degen}) for degeneracy to occur using

$$n_{\text{degen}} \approx \frac{1}{(\Delta \boldsymbol{r})^3} \approx 5 \left(\frac{m^*_{\text{ce}} kT}{h^2} \right)^{3/2}, \tag{3.3}$$

where h is the Plank's constant. For Ge, n_{degen} is in the order of $10^{18} - 10^{19}\,\text{cm}^{-3}$ at 300 K [1].

3.1.2 Temperature Dependent Energy Gap

The temperature dependent bandgap for intrinsic Ge can be fitted using a semi-empirical equation proposed by Varshni [41]

$$E_{\text{g}} = E_{\text{g}}(0) - \frac{\alpha T^2}{T + \beta} \tag{3.4}$$

where α and β are fitting parameters characteristic for a given semiconductor. β relates to the Debye temperature [23]. The values for $\alpha/(10^{-4}\,\text{eV K}^{-1})$ and β/K^{-1} for Ge have been obtained by Thurmond using graphical solutions , giving 4.774 ± 0.3 and 235 ± 40 for each constant respectively [40]. This results in a negative temperature coefficient for germanium's energy gap [39]. In addition, this formula assumes that changes in E_{g} is proportional to T at high temperatures and proportional to T^2 at low temperatures [40]. $E_{\text{g}}(0)$ approaches 0.743 eV at 0 K (figure 3.2).

A good fit is obtained using Varshni's equation for temperatures ranging from $200 - 450\,\text{K}$ [40]. (Data fitted was produced by MacFarlane *et. al.* by analyzing the absorption spectra of germanium using a spectrometer [18]). But for temperatures below 100 K onwards, Varshni's empirical model starts to deviate from experimental data [40]. This trend was observed for many semiconductors fitted using Varshni's equation, prompting many workers to produce other forms of empirical models that give better fits to E_{g} at lower temperatures [23, 27].

ATLAS by default implements Varshni's empirical equation to model the change in bandgap with temperature. Since this model accurately predicts the energy gap for germanium within the modulator's operating range ($250 - 350\,\text{K}$), no alteration was made [40]. For low temperature simulations (<100 K), Passler's model for energy dependent bandgap should be invoked instead [27]. Using four parameters instead of two,

[1] The electronic properties of semiconductor (e. g., bandgap and resistivity) no longer behave as expected in response to changes in electric field or temperature or any other disturbing influences.

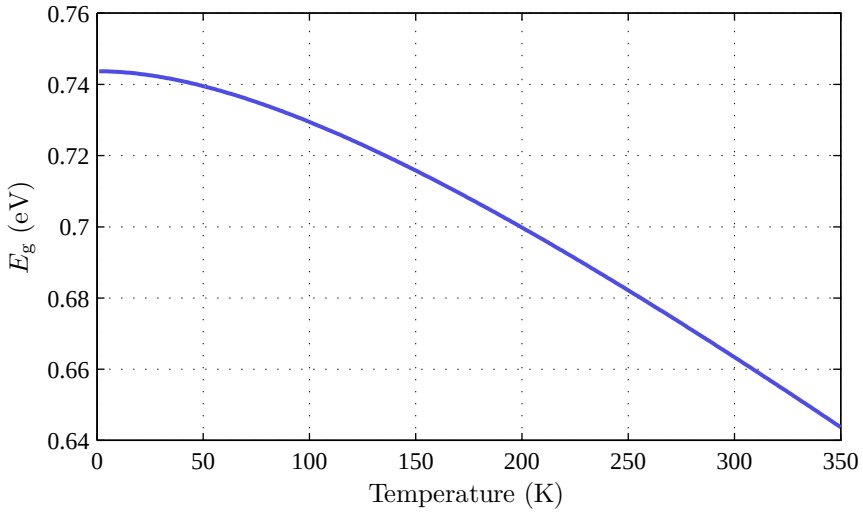


Figure 3.2: The energy gap of germanium as a function of temperature calculated using Eq. 3.4

this model accurately fits the energy gap of germanium down to 4 K [27]. In addition, parameters used also correlates well with specific properties of semiconductor [27]. (In Varshni's model, β, which is suppose to be related to the Debye temperature may, in certain cases, be negative! [23].)

3.2 Basic Theory on Carrier Statistics in Germanium

The density of thermally generated free electrons in undoped Ge is calculated using the expression [36]

$$n = \int_{E_C}^{\infty} N(E)\, F(E)\, \mathrm{d}E. \tag{3.5}$$

E_C is the energy at the lowest point of the conduction band (CB); $N(E)$ the total number of states available for occupancy; and $F(E)$ is the distribution function for carriers based on the classical Maxwell-Boltzmann distribution of molecular momenta [1]. $F(E)$ differs from Maxwell-Boltzmann's distribution because it incorporates Pauli's exclusion principle (no two carriers in the same system can have the same state), and the transitional probabilities of carriers in semiconductors [1]. Near the bottom of the conduction band, and at low-enough carrier densities and temperatures, $N(E)$ can be approximated by [36]

$$N(E) = M_C \frac{\sqrt{2}}{\pi^2} \frac{m_{de}^{3/2}(E - E_c)^{1/2}}{\hbar^3}. \tag{3.6}$$

M_C is the number of equivalent minima in the CB and m^*_{de} is the electron density of state effective mass given as $(m^*_l m^{*2}_t)^{1/3}$. (For Ge, the ratios of the longitudinal effective mass of electron to the electron rest mass, m^*_l/m_0, is 1.64; while the ratio between the effective transverse electron mass and the electron rest mass, m^*_t/m_0, is 0.082 [39]. Both ratio are valid only at 300 K, though studies by Julian showed that the effective mass for electrons and holes are hardly affected by the change in temperature [15]). Like the Maxwell-Boltzmann's distribution, $F(E)$ is only a function of the ambient temperature and the energy difference between states; indeed, it is independent of any other properties or arrangements of carriers in semiconductors under thermal equilibrium [1]. $F(E)$ can be represented using the Fermi-Dirac distribution function [12]

$$F(E) = \frac{1}{1 + \exp[(E - E_F)/kT]}, \tag{3.7}$$

where E_F is the Fermi energy level that results in an occupancy probability of 0.5 for states with energy equivalent to E_F [36]. Integrating Eq. 3.5 yields

$$n = N_C \frac{2}{\sqrt{\pi}} F_{1/2}\left(\frac{E_F - E_C}{kT}\right). \tag{3.8}$$

N_C is the conduction band effective density of states written as

$$N_C = 2\left(\frac{2\pi m^*_{de} kT}{h^2}\right)^{3/2} M_C; \tag{3.9}$$

while $F_{1/2}(x)$ is the Fermi-Dirac integral of order 1/2. By defining $\eta \equiv (E - E_C)/kT$ and $\eta_F \equiv (E_F - E_C)/kT$, the function $F_{1/2}\left(\frac{E_F - E_C}{kT}\right)$ in Eq. 3.8 can be expressed as [39]:

$$\int_0^\infty \frac{\eta^{1/2}}{1 + \exp(\eta - \eta_F)} \equiv F_{1/2}(\eta_F). \tag{3.10}$$

If $\eta_F < -1$, a condition satisfied in non-degenerate semiconductors, Eq. 3.10 can then be simplified to

$$\frac{\sqrt{\pi}}{2} \exp\left(-\frac{E_C - E_F}{kT}\right). \tag{3.11}$$

In a similar way, the density of holes in Ge can be obtained via the equation

$$p = N_V \frac{2}{\sqrt{\pi}} F_{1/2}\left(\frac{E_V - E_F}{kT}\right), \tag{3.12}$$

where E_V is the energy at the top edge of the VB. The hole effective density of states is given as

$$N_V = 2\left(\frac{2\pi m^*_{dh} kT}{h^2}\right)^{3/2}. \tag{3.13}$$

m^*_{dh} is the hole density of state effective mass given as $({m^*_{lh}}^{3/2} + {m^*_{hh}}^{3/2})^{2/3}$ (See section 2.2.2). For Ge, $m^*_{lh}/m_0 = 0.04$ and $m^*_{hh}/m_0 = 0.28$ at 300 K [39]. If E_F

is several kT above E_V (i. e., $\eta_F < -1$), the Fermi-Dirac integral for calculating the density of holes in Eq. 3.12 can reduced in a similar manner to Eq. 3.11. Hence, for non-degenerate semiconductors, the electron and hole density can be written as:

$$n = N_C \exp\left(\frac{E_F - E_C}{kT}\right) \tag{3.14}$$

$$p = N_V \exp\left(\frac{E_V - E_F}{kT}\right) \tag{3.15}$$

respectively. Finally, by defining the Fermi energy level for intrinsic semiconductor (E_i) as [37]

$$\frac{E_C + E_V}{2} + \frac{kT}{2} \ln\left(\frac{N_V}{N_C}\right) = E_F = E_i; \quad E_C + E_V = E_g(T), \tag{3.16}$$

where E_g is the energy gap discussed in section 3.1, we now have the complete set of equations needed to compute the intrinsic carrier concentration for Ge as a function of temperature. Alternatively, one could also multiply n and p to form

$$N_C N_V \exp\left(-\frac{E_g}{kT}\right) = (n_i)^2. \tag{3.17}$$

n_i is the intrinsic carrier concentration that is a good approximation for both the concentration of electrons and holes in semiconductors [1].

Together with Varshni's temperature dependent bandgap model, Eq. 3.7 to 3.16 are utilized by ATLAS for computing the concentration of carriers in semiconductors [35]. In non-degenerate semiconductors, the product np is pegged to the square of n_i (i. e., $np = {n_i}^2$), and is known as the mass-action law [12]. As a consequence, donors (electron contributor) or acceptors (hole contributor) can be added into semiconductor to produce n or p-type samples respectively. In the former, the majority of carriers consist of electrons, while for the latter, holes.

The introduction of dopants in semiconductor shifts the fermi energy level away from the intrinsic energy level E_i [39]. The following equations can be used to estimate the shift in E_F [39]

$$N_D = n_i \exp\left(\frac{E_F - E_i}{kT}\right) \tag{3.18}$$

$$N_A = n_i \exp\left(\frac{E_i - E_F}{kT}\right), \tag{3.19}$$

where N_D and N_A are the concentration of donor and acceptor ions respectively. The change in E_F as a function of temperature and doping concentration are evaluated in figure 3.3 using equations 3.4, 3.16, 3.17, 3.18 and 3.19, all of which are implemented in ATLAS.

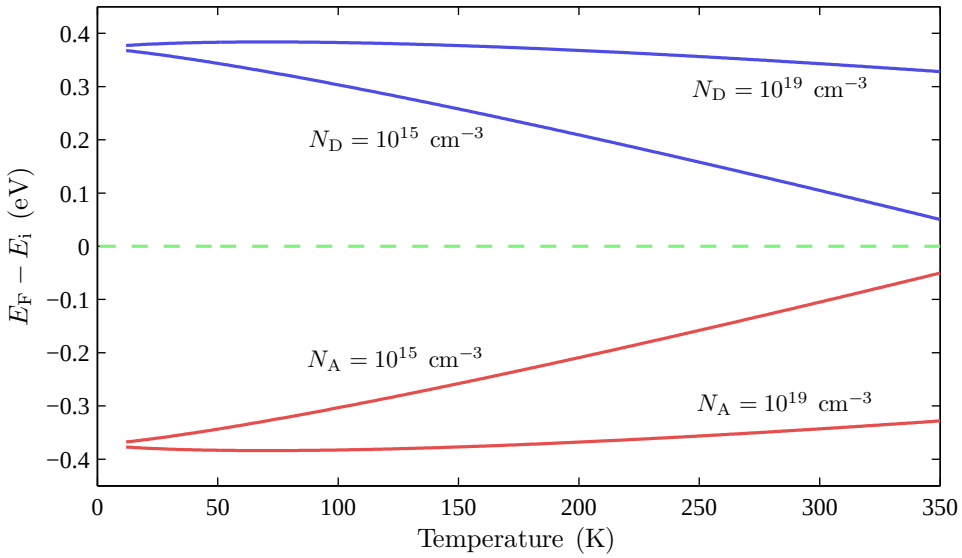


Figure 3.3: Change in the Fermi level as a function of temperature and dopant concentration.

Finally, several user definable parameters are available in ATLAS for computing the intrinsic concentration of carriers. If default settings are not used, N_C and N_V can be entered directly (equations 3.9 and 3.13), or the conductivity effective mass of carriers (m^*_{de} and m^*_{dh}) can be specified instead. Results from manually entering the aforementioned parameters were compared with ones using ATLAS's default settings for Ge. All methods produced very similar results within the temperature range of $223 - 350\,\mathrm{K}$. The default settings were therefore used for simulations.

3.2.1 Fermi-Dirac and Boltzmann Statistics

The Fermi-Dirac distribution function, $F(E)$, describes the probability of allowed energy states being occupied by carriers [1]. It is used in Eq. 3.5 to evaluate the intrinsic concentration of electron and holes in semiconductors. The downside of utilizing this distribution is that the resulting equations (Eq. 3.5 and 3.12) contains the function $F_{1/2}(\eta_F)$, which cannot be solved easily. Although inaccurate approximations are available, ATLAS uses the Rational Chebyshev approximation (a power series derived from the Chebyshev polynomial given as: $F_{1/2}(\eta_F) = y\left(1 - \frac{y}{2^{1/2+1}} + \frac{y^2}{3^{1/2+1}} - \frac{y^3}{4^{1/2+1}} + \cdots\right)$; $\eta_F > -1$ and $y = \mathrm{e}^{\eta_F}$) for evaluating $F_{1/2}(\eta_F)$, giving results close to exact values [19, 35]. This method is, of course, computationally intensive. Coupled with the relatively large size of the modulator (ATLAS is predominantly used for simulating micron sized devices while the modulator has a volume of $\geq 200\,\mathrm{mm}^3$), a considerable amount

of time will be needed in predicting and optimizing the electrical performance of the modulator. As an example, a 2 mm thick modulator with only basic physical models implemented takes 55 minutes for a transient simulation to complete (the use of a more powerful server with opteron processors installed subsequently reduced the time required by a factor $5 - 10$).

Fortunately, a simpler form of distribution known as Boltzmann's statistics can be employed instead—provided that samples are non-degenerate. This distribution is given as [39]:

$$F(E) = \exp\left(\frac{E_{\mathrm{F}} - E}{kT}\right), \tag{3.20}$$

which only consists of a single exponential term. When used, simulation run times can roughly be reduced by half the original time required. Therefore, it is important that the suitable distribution is invoked to prevent the wastage of computing time. The choice between Fermi-Dirac and Boltzmann's statistic depend on whether degenerate regions are present in structures to be simulated. We have, in the previous section, dealt with degeneracy in terms of dopant and carrier concentrations. A more accessible method for evaluating the state of the semiconductor is by comparing Eq. 3.10 and 3.20. With properties of Ge assumed, both distributions are plotted in figure 3.4 as a function of η_{F}. Boltzmann's statistic is valid so long as its values are similar to its counterpart—

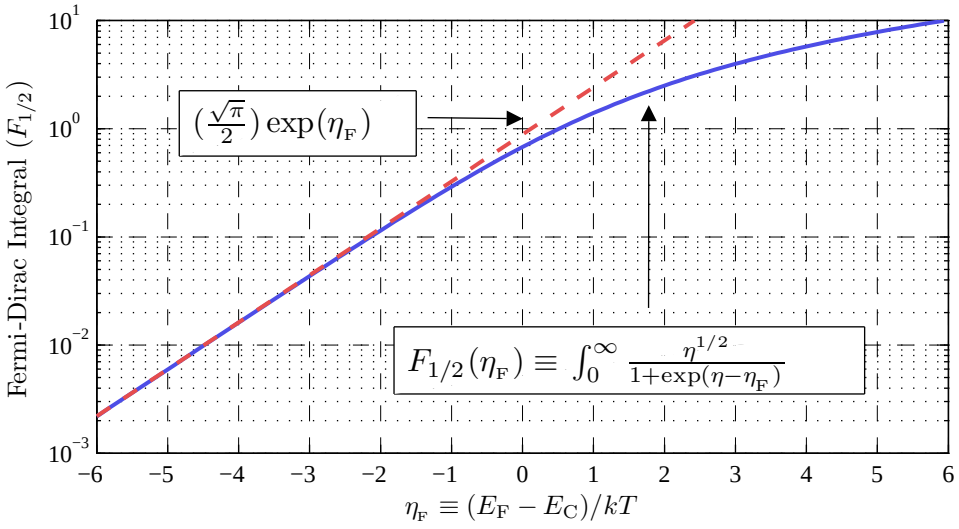


Figure 3.4: Fermi-Dirac integral as a function of η_{F}. For η_{F}>-1, the approximate equation using Boltzmann's statistic (dotted line) deviates from the Fermi-Dirac Integral function.

that is, in regions where η_{F} is less than -1 (figure 3.4). For $\eta_{\mathrm{F}} > -1$, Boltzmann's

statistic deviates from Fermi-Dirac distribution, rendering it inappropriate for use in simulation. The conditions for degeneracy can thus be be considered in terms of η_F. In addition, the dopant concentration in which Boltzmann's statics is valid can be evaluated as well. The results are shown in figure 3.5. The erroneous use of Boltzmann's

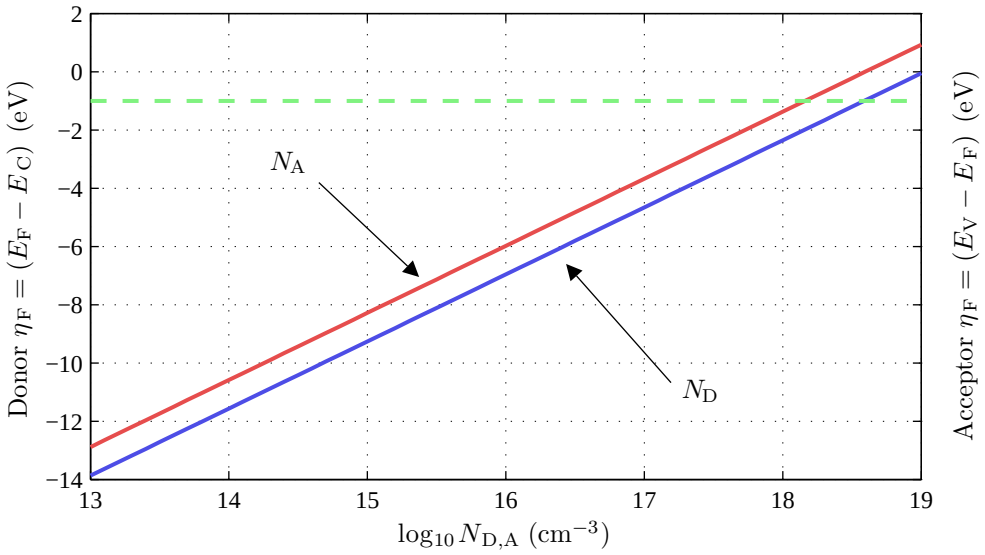


Figure 3.5: Degeneracy occurs for η_F exceeding -1. (N_D and N_A are the concentrations of donor and acceptor respectively.)

statistic will result in the generation of a higher number of carriers in degenerate semiconductors (Eq. 3.5 and figure 3.4), which may degrade the performance of the modulator due to unwanted minority carrier injection at forward bias (section 4.2.1). This will worsen the performance of the modulator.

Except for extremely shallow degenerate regions beneath the aperture and the electrodes of the modulator (depths <50 nm measured using secondary ion mass spectrometry), the bulk of the diode is made up of intrinsic Ge (i. e., > 99.999% of the total volume). Therefore, Boltzmann's distribution was invoked in ATLAS for computing the concentration of carriers in Ge. Doing so makes subsequent calculations in ATLAS *much* faster [35]. As a precaution, Fermi-Dirac and Boltzmann's distribution was compared in simulation using the default modulator structure. Results obtained yielded no noticeable difference within the scale that is of interest to us. (The concentration of injected carriers is roughly 3 orders of magnitude larger compared to the intrinsic carrier concentration. Thus any changes brought about by the thin degenerate layers in the modulator are negligible. This assumption will not be true if a thick layer of highly doped region is present because of increased carrier recombination and changes

to injection efficiency.)

3.2.2 Incomplete Ionization of Impurities

Carriers from donor and acceptor atoms must first be liberated before they contribute to the electrical properties of the host semiconductor (i. e., to be electrically "active"). Based on the simple hydrogen atom model, the following equation can be used as a rough approximate for the ionization energy of donor electrons in semiconductors [37]:

$$E_{\mathrm{C}} - E_{\mathrm{D}} = \left(\frac{\epsilon_0}{\epsilon_{\mathrm{s}}}\right)^2 \left(\frac{m^*_{\mathrm{ce}}}{m_0}\right) E_{\mathrm{H}} \tag{3.21}$$

where E_{H} is the ionization energy of a hydrogen atom ($\approx 13.6\,\mathrm{meV}$); ϵ_0 and ϵ_{s} the vacuum and semiconductor permittivity respectively; and E_{D} is the donor energy level. In Ge, the donor ionization energy comes to around 30 and 20 meV for donors and acceptors respectively. At room temperature, one could thus expect a good number of impurities to be ionized since the thermal energy at 300 K is $\approx 26\,\mathrm{meV}$.

However, at doping levels above N_{C} or N_{V}, Ge becomes degenerate, and in many respects, behaves like an ordinary metal (e. g., monotonic increase in resistivity with temperature, carrier densities that hardly changes between 300 and 11 K etc) [34]. As a consequence, the ratio of ionized impurities in degenerate Ge are, for a very broad range of temperatures, almost the same. Various workers have confirmed this attribute for temperatures down to 11 K [1, 4, 34, 36].

On the other hand, impurity ionization in non-degenerate Ge is a strong function of temperature [1, 39]. The ratio of ionized donors and acceptors in Ge, denoted by N_{D}^+ and N_{A}^- respectively, can be written using the following equations:

$$N_{\mathrm{D}}^+ = \frac{N_{\mathrm{D}}}{1 + g_{\mathrm{D}} \exp[E_{\mathrm{F}} - (E_{\mathrm{C}} - E_{\mathrm{D}})/kT]} \tag{3.22}$$

$$N_{\mathrm{A}}^+ = \frac{N_{\mathrm{A}}}{1 + g_{\mathrm{A}} \exp[(E_{\mathrm{A}} - E_{\mathrm{V}} - E_{\mathrm{F}})/kT]} \tag{3.23}$$

where E_{D} and E_{A} are the donor and acceptor energy levels; g_{D} is the ground-state degeneracy of donor impurity level with value of 2, because an electron of either spin can reside in a donor energy state; while g_{A} corresponds to the ground-state degeneracy for acceptor levels. In semiconductors such as Si and Ge, because impurity energy levels are doubly degenerate at Γ ($k = 0$), the value of g_{A} is twice the value of g_{D} [39].

Equations 3.22 and 3.23 are activated in ATLAS to account for changes in the concentration of ionized impurities at low temperatures. Appropriate degeneracy factors

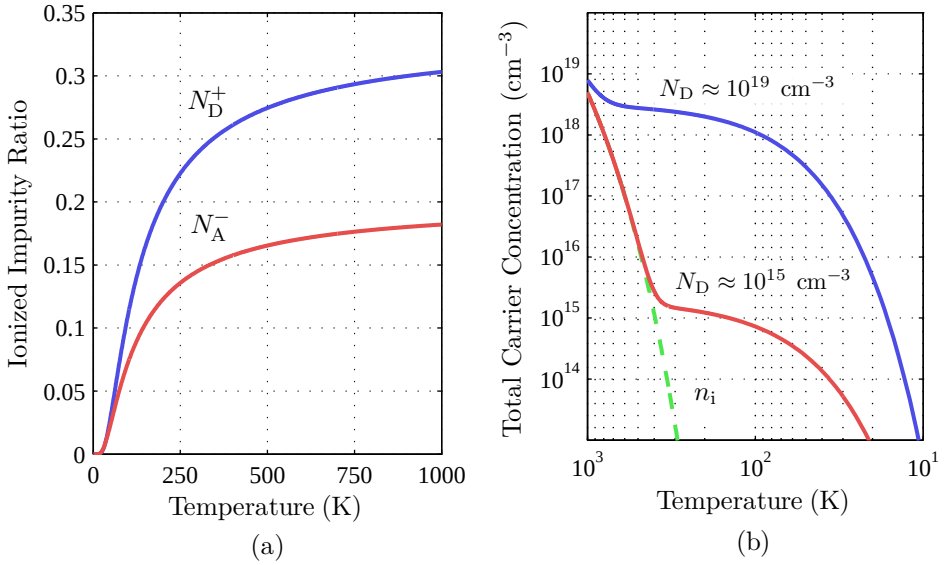


Figure 3.6: (a) the fraction of ionized donors and acceptors in non-degenerate Ge (for densities below N_C or N_V); (b) the total concentration of free carriers in Ge as a function of temperature and donor concentration. The concentration of carriers due to ionized donors hardly changes within the modulator's operating range of 300 $\pm$ 50K (compare with the intrinsic carrier concentration n_i, where free carriers are formed due to thermal agitation alone).

can be set using the parameter `GCB` for g_D, and `GVB` for g_A. Because Boron and Phosphorus were used as acceptors and donors in Ge (via ion implantation), corresponding values of $(E_V - E_A) = 0.01\,\text{eV}$ and $(E_C - E_D) = 0.012\,\text{eV}$ were used [37, 38].

At low temperatures (<100 K), the effects of impurity ionization are important in analyzing the electrical and optical performance of the modulator due to the following reasons:

- The injection efficiency of carriers into the modulator, which in turn influences the rate of carrier diffusion, is strongly dependent on the concentration of donors and acceptors beneath the electrode (section 4.2.1). This directly affects the performance of the modulator.

- The concentration of donors and acceptors implanted at doses of $8 \times 10^{14}\,\text{cm}^{-2}$ and $5 \times 10^{14}\,\text{cm}^{-2}$ respectively are sufficient to reduce the absolute transmission by 0.05 for $\lambda = 10.59$ at 300 K (Eq. 3.22 -3.23 and figure 2.9). Therefore, for low temperature optical measurements (section 7.5.1), the change in the fraction of ionized impurity must be considered. (Based on Eq. 3.22 and 3.23, the concentration of ionized donors is predicted to drop by $> 25\%$ when the temperature

drops from 300 K to 77 K.)

Values projected in figure 3.6 show that between 250 and 350 K, the fraction of ionized impurities do not vary much. Thus within that temperature range, the effects of carrier freeze-out are not significant to the performance of the modulator. But for low temperature experiments, changes to the electrical and optical properties of the modulator must be carefully considered for proper evaluation of its performance.

3.3 Carrier Mobility

For PiN diodes operated under forward bias, carrier mobility plays an important role in determining the electronic properties of injected carriers. The concentration of electrons and holes across the diode, and its current voltage (I-V) characteristics are examples of parameters that are dependent on carrier mobility 4.2.1.

Early studies on the mobilities of electrons and holes in Ge were often deduced from the conductivity and Hall effect measurements [21]. In many of these works, authors have emphasized that considerable amount of simplifications were used in deriving the values for carrier mobilities [4, 10, 21]. Unfortunately, these data are the only sources available within the range of dopant concentration (10^{13} to $10^{19}\,\mathrm{cm}^{-3}$) and temperature change ($223 - 323\,\mathrm{K}$) that are of interest to us—evident in their use by many texts on semiconductor physics [1, 20, 36, 37]. The scarcity of data on carrier mobility in Ge is due to the shift of interest in the electronics community towards the use of silicon (Si), propelled by the need for solid state amplifiers with higher temperature tolerance, and the discovery of SiO_2 formation on Si wafers that proved to be tremendously useful in planar process fabrication [11, 29]. The quality of samples used in early studies is also questionable due to the inclusion of contaminants such as Cu. In n-type samples for example, if dopant concentration exceeds $10^{16}\,\mathrm{cm}^{-3}$, it is normal for Cu to be present in densities around the order of $10^{15}\,\mathrm{cm}^{-3}$ [10, 22]. Since substitutional Cu in Ge acts as a trivalent acceptor, the concentration of donors (N_D) should be more than three times the concentration of Cu for the sample to be rendered n-type [22]. Needless to say, the analysis of compensated samples involves additional simplifications in estimating the contribution of impurity scattering towards carrier mobility [10].

In fitting the mobility data of carriers for use in ATLAS, sources dating back to the fifties and early sixties—though questionable—were used directly without further validation. This approach is justified for the following reasons: a) to the best of our knowl-

edge, no alternative publication on carrier mobility were found, and b) samples used in sources cited were homogeneously doped single crystals with good crystallinity [4]. Whereas in the modulator, dopants were forcefully implanted into Ge crystal at high doses, which leads to the formation of inactive impurity clusters [30], and damages to the lattice structure [39]. Since carrier mobility is very sensitive to the arrangement of donors and acceptors in the host crystal [1], the mobility of carriers in implanted samples would certainly differ from ones used in early references. Nonetheless, due to insufficient data, we have assumed—without justification given—that differences between the two are not significant. These reasons negate the need for careful analysis of early mobility data.

In the following sections, different deflecting or 'scattering' forces that act to change the *drift* mobility of carriers (i. e., carriers driven by a potential field) are briefly presented. Experimentally determined data on carrier mobility are then fitted using empirical models available in ATLAS.

3.3.1 Lattice Scattering

Lattice or thermal scattering is by far the most important mechanism that alters the mobility of carriers in non-degenerate semiconductors [1]. It has been analyzed using two points of view, whereby distortions in the lattice due to thermal vibrations can be regarded as: (a) deformation potentials in the crystal (lattice vibrations treated as a waves), or (b) the movement of quantized phonons in the crystal [5]. In the former, vibrational waves travelling through the lattice superimpose with the periodic potential of the crystal, creating potential ripples that scatters carriers; while in the latter, scattering is caused by the interactions between electrons and phonons. According to Dunlap, both methods yield essentially the same result in spite of the difference in approach [5]. We have adopted the second viewpoint, which predicts the dependence of lattice scattering based on two modes of phonons, described using the following equations:

$$\mu_{\mathrm{ac}} = \mathrm{A}T^{-1.5} \tag{3.24}$$

$$\mu_{\mathrm{op}} = \mathrm{B}T^{-0.5}(\mathrm{e}^{\Theta/T-1}). \tag{3.25}$$

μ_{ac} and μ_{op} are the mobilities of carriers scattered by acoustic and optical phonons respectively. A, B and Θ are constants. Morin gives the solution for Eq. 3.24 and 3.25

as [21]:

$$\mu_{\mathrm{ac}} = 2.4 \times 10^{7} T^{-1.5} \tag{3.26}$$

$$\mu_{\mathrm{op}} = 7.8 \times 10^{4} T^{-0.5} (\mathrm{e}^{520/T-1}) \tag{3.27}$$

for electrons, and

$$\mu_{\mathrm{ac}} = 2.5 \times 10^{8} T^{-2.0} \tag{3.28}$$

$$\mu_{\mathrm{op}} = 1.9 \times 10^{4} T^{-0.5} (\mathrm{e}^{520/T-1}) \tag{3.29}$$

for holes. With μ_{ac} and μ_{op} defined, the mobility of carriers solely due to lattice scattering can be approximated using [21, 22]

$$\frac{1}{\mu_{\mathrm{L}}} = \frac{1}{\mu_{\mathrm{ac}}} + \frac{1}{\mu_{\mathrm{op}}}. \tag{3.30}$$

The mobility of electrons and holes under the influence of thermal scattering can thus be simplified to [21, 22]

$$\mu_{\mathrm{L_n}} = (4.9 \times 10^{7})\, T^{-1.66} \tag{3.31}$$

$$\mu_{\mathrm{L_p}} = (1.05 \times 10^{9})\, T^{-2.33}. \tag{3.32}$$

Though the derivation of Eq. 3.31 and 3.32 are highly speculative (they depart from the traditional $T^{-1.5}$ law predicted by theory, which excludes the scattering effects of optical phonons [39]), both agree well with results obtained experimentally [21]. Subsequent publications from various workers have confirmed the $T^{-1.66}$ and $T^{-2.33}$ power dependency for the mobility of electrons and holes respectively [37].

3.3.2 Impurity Scattering

Impurity scattering (IS) involves the direct electrostatic interaction between ionized dopants with moving carriers that result in scattering [5]. Dopants will either attract or repel carriers according to their ionization state and the charge sign of carriers. The efficiency of this scattering mechanism depends on the vector of moving carriers with respect to bound ions in the lattice [1]. In modelling the dependence of IS mobility, the Conwell-Weisskopf formula was widely used before it was superseded by the Brooks-Herring formula (BH). The latter takes into account screening effects by carriers, which result in the increase in carrier mobility at high carrier concentrations [10]. It is given as [33]:

$$\mu_{\mathrm{I}} = \frac{2^{7/2} (4\pi \epsilon_{\mathrm{s}})^2 (kT)^{3/2}}{\pi^{3/2} Z^2 q^3 (m^*)^{1/2} N_{\mathrm{I}}} \left[\frac{1}{\ln(1+\beta_{\mathrm{BH}}) - \beta_{\mathrm{BH}}/(1+\beta_{\mathrm{BH}})} \right], \tag{3.33}$$

where Z is the charge of impurity centre, N_I the concentration of dopants, m^* is the conductivity effective mass of carriers (we have assumed $m_\mathrm{n}^* = 0.12m_0$ and $m_\mathrm{p}^* = 0.28m_0$), and ϵ_s the semiconductor permittivity. β_BH, a function with modifications from Brooks and Herring [33], is written as:

$$\beta_\mathrm{BH} = \frac{2\,m^*\,\epsilon_\mathrm{s}\,(kT)^2}{\pi\,h^2\,q^2\,n}. \tag{3.34}$$

Assuming singly charged dopants ($Z = \pm 1$), the carrier mobility affected by lattice and impurity scattering in Ge are compared in figure 3.7. At 300 K, lattice scattering

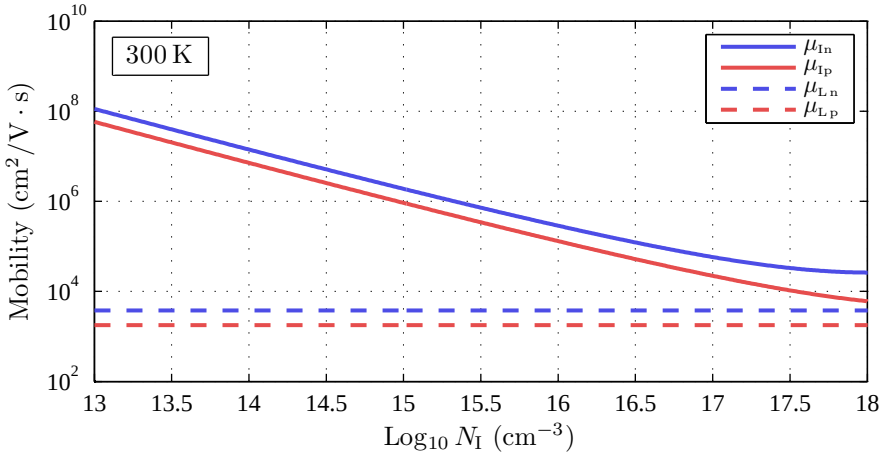


Figure 3.7: Comparison between carrier mobility due to lattice (μ_L) and impurity (μ_I) scattering . The slope of μ_I is reduced when carrier concentration exceeds $5 \times 10^{17}\,\mathrm{cm}^{-3}$ due to screening. (Subscripts 'n' and 'p' denote electrons and holes respectively.)

remains the dominant scattering mechanism even when dopants are present at very high concentrations.

Besides ionized dopants, neutral unionized donors and acceptors too contribute toward the scattering of carriers. This is known as neutral impurity scattering (NIS) whereby localized strains in the crystal due to the presence of impurities act to scatter carriers [1, 5]. The close analogy between neutral dopants and a hydrogen atom have led Pearson and Bardeen to treat the subject of NIS based on the scattering of slow electrons by hydrogen atoms [26]. Using similar approach, a more exact solution was proposed by Erginsoy, giving the dependence of electron mobility as [6]:

$$\mu_\mathrm{n} = \frac{1}{20}\frac{m_\mathrm{n}^* q^3}{N_\mathrm{N}\epsilon_\mathrm{r}\hbar^3}, \tag{3.35}$$

where N_N and ϵ_r are the concentration of neutral dopants and the semiconductor relative permittivity respectively.

Although NIS is relatively weak and is usually obscured by other scattering mechanisms [4], its effective scattering cross section becomes comparable to ionized impurity scattering at temperatures in the order of 100 K [6]. The contribution of NIS is, of course, included in the Hall data. This enables the effective carrier mobility to be extracted for use in simulations directly without further modifications to carrier mobility.

3.3.3 Carrier-carrier Scattering

Carrier-carrier scattering (CCS) is similar to impurity scattering except that carriers deflect each other in passing about their common centre of mass [1]. CCS can be divided into two categories: collisions between electrons (or between holes), and between electrons *and* holes [36].

Experiments involving non-degenerate semiconductor devices have shown that inter electron scattering have negligible effects on carrier mobility, as one may deduce from metals where such collisions hardly affect the electrical resistance of metals [1, 36]. Physically, it mostly alters the path of individual electrons, whilst the momentum of the entire group, driven by the same electric field, is conserved. Inter electron scattering does however contribute to the "viscosity" of carriers (analogues to the frictional flow of gas molecules), which is dependent upon the boundary condition at the surface not addressed in Ohm's law [8]. This could be important for fine structured, planar devices (e. g., CMOS transistors or photoconductors), but not for large bulky devices such as ours [5, 36].

The contribution of electron-hole scattering towards carrier mobility is generally small, and only a limited amount of experimental evidence have been published on its magnitude [1, 9, 36]. Though information on this type of scattering is scant, it is generally observed that when carriers are present in large enough quantities, CCS becomes strong enough to compete even with lattice scattering—the dominant scattering mechanism in Ge [36]. Using the modified impurity scattering formula by Conwell and Weisskopf, and Brooks and Herring, the electron-hole scattering mobility can be evaluated using [21]

$$\mu_{\mathrm{eh}} = \frac{2^{7/2}\,\epsilon_{\mathrm{s}}^2\,(kT)^{3/2}}{\pi^{3/2}\,(m_{\mathrm{n}} m_{\mathrm{p}}/[m_{\mathrm{n}}+m_{\mathrm{p}}])^{1/2}} \left\{ q^3\,(np)^{1/2}\,[\ln(1+B) - B/(1+B)] \right\}, \tag{3.36}$$

where the parameter B is given as

$$B = \frac{6\,\epsilon_{\mathrm{s}}\,(m_{\mathrm{n}} m_{\mathrm{p}}/[m_{\mathrm{n}}+m_{\mathrm{p}}])(kT)^2}{(np)^{1/2}\hbar^2 q^2}. \tag{3.37}$$

In intrinsic Ge, the crossover point where electron-hole scattering becomes dominant for different temperatures is shown in figure 3.8. Between 223 and 323 K, excess carriers

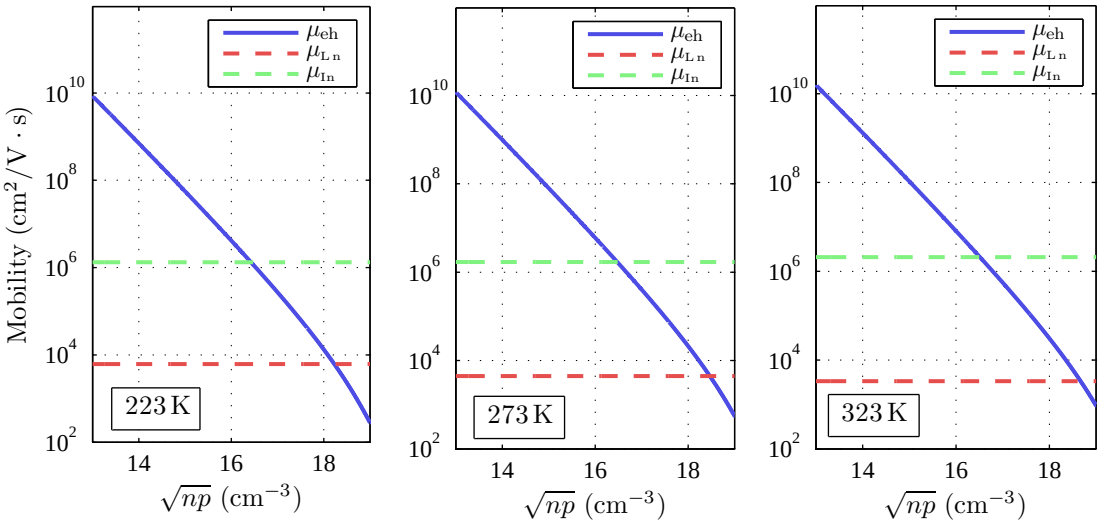


Figure 3.8: Comparison between electron-hole scattering mobility with μ_{Ln} and μ_{In}. The concentration of ionized donors were set to $10^{11}\,cm^{-13}$. (Excess carriers are introduced via electrical injection)

must exceed $10^{18}\,cm^{-3}$ for electron-hole scattering to surpass lattice scattering.

At a fixed carrier concentration of $5 \times 10^{16}\,cm^{-3}$ (the concentration required to induce $> 95\%$ IR modulation at $10.59\,\mu m$ by the modulator at 300 K), only at temperatures below 50 K is CCS more significant as compared to lattice scattering. This is depicted in figure 3.9.

Several models for carrier-carrier scattering are available in ATLAS. Nevertheless, they were not invoked because its effect is negligible within the operating conditions of the modulator, where the concentration of carriers does not exceed $10^{18}\,cm^{-3}$ [1, 4, 21].

Having discussed scattering mechanisms relevant to the performance of the modulator, the mobility of carriers experimentally determined from conductivity and Hall effect measurements are fitted using empirical equations in the next section.

3.3.4 Temperature and Dopant Dependent Carrier Mobility

The temperature and doping dependent model for carrier mobility is based on the work of Caughey and Thomas [35]. Because it contains several parameters which have to be hand fitted simultaneously, the doping dependence of carrier mobility is first

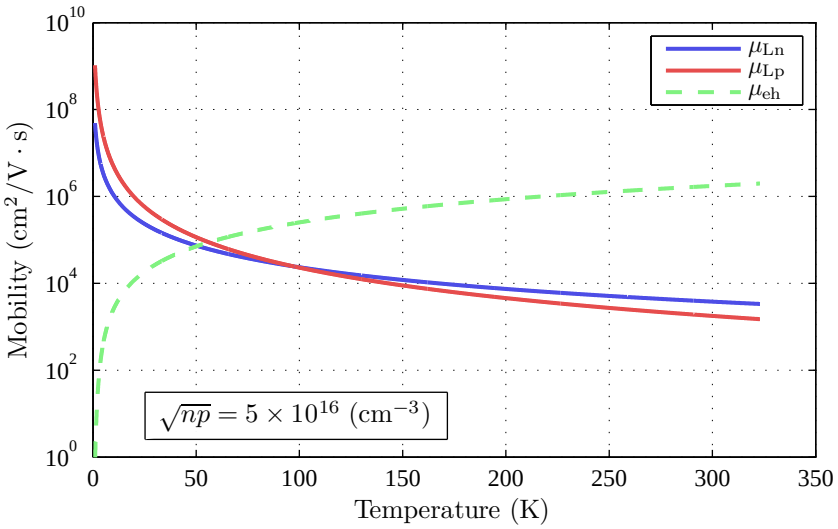


Figure 3.9: Comparison between lattice and electron-hole scattering at fixed excess carrier concentration. (Excess carriers are introduced via electrical injection.)

considered using the method proposed by Caughey and Thomas [3]. Data published by Moll, being the most recent, were used for this purpose [20].

The plots of measured mobility versus the logarithm of doping density can be fitted using the hyperbolic tangent function in the form

$$\mu_{\mathrm{I}} = \frac{\mu_{\max} - \mu_{\min}}{1 + (N_{\mathrm{I}}/N_{\mathrm{ref}})^{\alpha}} + \mu_{\min}, \tag{3.38}$$

where μ and N_{I} are the mobility and dopant concentration respectively. By rearranging Eq. 3.38 to

$$(N_{\mathrm{I}}/N_{\mathrm{ref}})^{\alpha} = (\mu_{\max} - \mu)/(\mu - \mu_{\min}), \tag{3.39}$$

the values for $\mu_{\min}$ and $\mu_{\max}$ were chosen, by trial and error, to give the best fit of a straight line for Eq. 3.39. α and N_{ref} were then obtained from the slope and unity intercept of the straight line. The fitted mobility data for electrons and holes in Ge are shown in figure 3.10 and 3.11.

Temperature dependence can then be incorporated into Eq. 3.38 using [35]

$$\mu_{\mathrm{I}} = \frac{\mu_{\max}\left(\frac{T}{300}\right)^{\texttt{BETAN.CAUG}} - \mu_{\min}\left(\frac{T}{300}\right)^{\texttt{ALPHAN.CAUG}}}{1 + \left(\frac{T}{300}\right)^{\texttt{GAMMAN.CAUG}}(N_{\mathrm{I}}/N_{\mathrm{ref}})^{\alpha}} + \mu_{\min}\left(\frac{T}{300}\right)^{\texttt{ALPHAN.CAUG}}. \tag{3.40}$$

As discussed in section 3.3.1, `BETAN.CAUG` has a dependence of -1.66 and -2.33 for electrons and holes respectively. `ALPHAN.CAUG` was set to null as suggested in the ATLAS manual [35], and `GAMMAN.CAUG` was determined by trail and error.

Electron and hole mobility by Debye and Golikova were then fitted using Eq. 3.40 [4, 10], and the results are shown in figure 3.12 and 3.13. Only the curve for electron

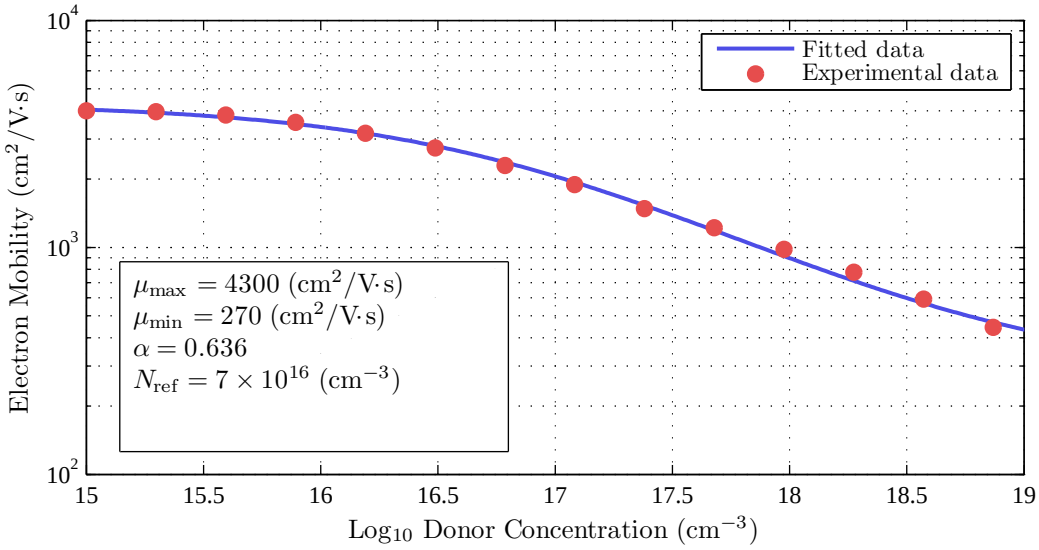


Figure 3.10: Fitted mobility data for Electrons in Ge

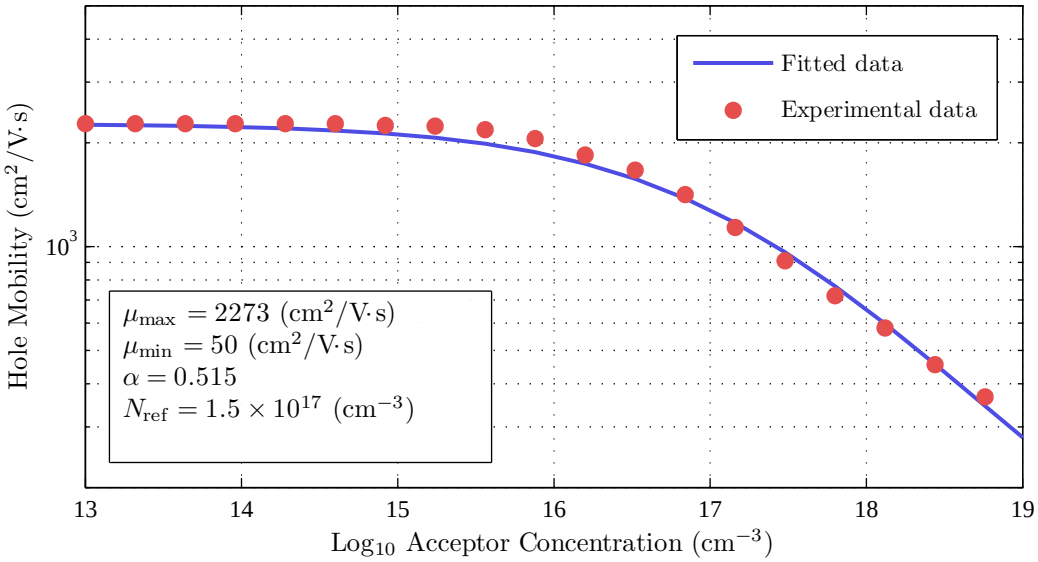


Figure 3.11: Fitted mobility data for Holes in Ge

mobility at donor concentration of $1.7 \times 10^{15}\,\mathrm{cm}^{-3}$ deviate from experimental data (figure 3.12). It is hard to give a precise reason for this behaviour, since good fits were achieved for intrinsic Ge and for mobility at higher doping concentration. The quality of sample used, or complications in the Hall effect involved in mobility measurements could be factors contributing to the observed behaviour [1]. Although more recent data from Fistul are available [7], it is poorly written with no reference to the type of mobility measured. (In conductivity and Hall effect measurements, two different mobility are usually given. They are the Hall and drift mobility. Fistul did not state which of the two he measured. He did however compared his data with that of Debye, which were presented using Hall mobility—but one could easily convert from one form to the

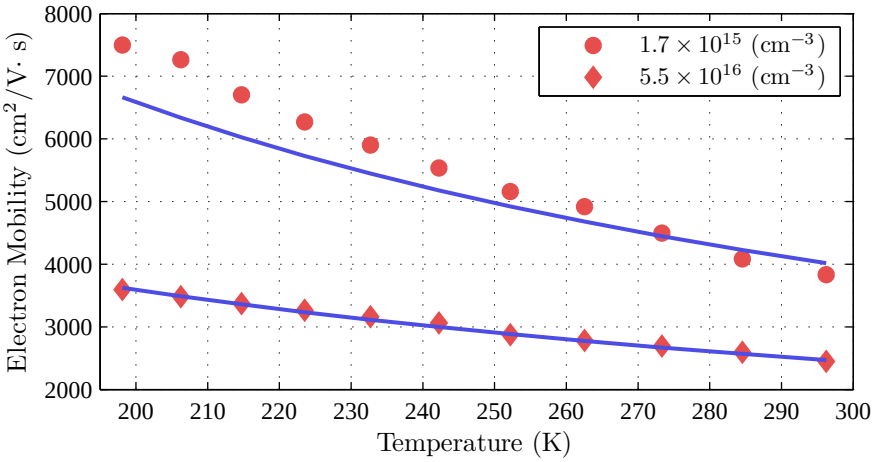


Figure 3.12: Temperature and doping dependent electron mobility data for two different donor concentrations. α of -0.7 was used.

other). Also, whether it is really Hall or drift mobility Fistul provided, the mobility of electrons are significantly underestimated in his paper. The Hall mobility relates to the drift mobility by the equation

$$\mu_{\mathrm{H}} = r_{\mathrm{H}}\,\mu \tag{3.41}$$

In general, r_{H} lies in the range of $1-2$, and is a strong function of carrier scattering [10, 37]. Since Fistul gives no information on values of r_{H} used, and his data for lightly doped samples are $\approx 25\%$ less compared to other literature (assuming r_{H} of 1) [20, 37], have discouraged us from using his data.

3.3.5 Low Field Mobility Approximation

In modelling the mobility of carriers in Ge under the influence of electric fields, it can be divided into (a) low field behaviour or (b) high field behaviour. The former, valid for fields below 1000 V/cm, assumes that the movement of carriers is strongly dependent upon phonon and impurity scattering—both of which act to decrease carrier mobility[14, 28]. On the other hand, if the electric field exceeds > 1000 V/cm, carriers then gain sufficient energy to take part in a wider range of scattering processes, resulting in smaller velocity increase with increasing field. The velocity of carriers eventually saturates to a constant value at very high values of electric fields. This nonlinear change in carrier velocity is modelled using high field approximation, and is primarily a function of lattice scattering [35]. (The drift mobility of carriers in semiconductors at high electric fields are up to date an open problem, and many aspects are not fully

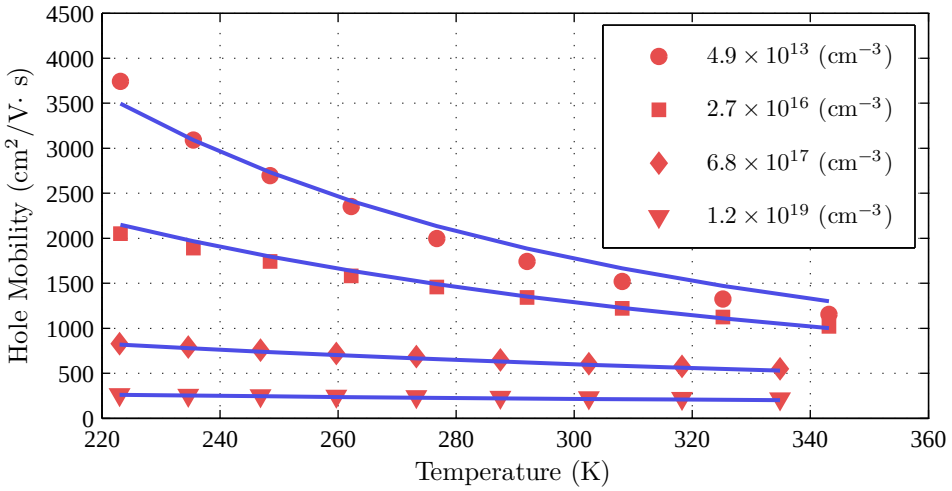


Figure 3.13: Temperature and doping dependent hole mobility for four different acceptor concentrations. α of -1.7 was used.

understood [24]. Smith suggested that other forms of unknown scattering mechanisms may be present [36]. Both ionized impurity scattering and a mixture of this and lattice scattering have been considered by Gunn, but he was unable to account for variations at moderate electric fields [13]. ATLAS fits the mobility of carriers at high electric field using semi-empirical equations. Alternatively, by incorporating the warping and non-parabolicity effects of the conduction and valence band (caused by the change in electric fields) to estimate the mobility of carriers, satisfactory agreement between theory and experimental results were obtained by Jacoboni *et. el.* and Reggiani *et. el.* [14, 28]. Their work thus differs from the mainstream view, which attribute the non-linear rise in carrier mobility at high electric fields to some form of carrier scattering) Since the modulator requires no more than 1.5 V of forward voltage during 'on' state, the average electric fields in the modulator will not exceed 15 V/cm—roughly two orders of magnitude below the requirements for high field approximation. Therefore, low field approximation was used in modelling the electrical behaviour of the modulator.

3.4 Metal-Semiconductor Contacts

Current is supplied to the modulator through metal electrodes evaporated on its surfaces. To ensure low power dissipation, it is vital that electrical resistance on contacts formed be kept low. Low resistance contacts are known as ohmic contacts. They have linear or quasi-linear I-V characteristics, and should able to supply the current required by respective devices without incurring large voltage drop across the contacts [36].

In producing low resistance contact, three factors that contribute to the departure from ohmicity must be considered. They are: difference in work function, surface states, and resistive interfacial layers [32, 39]. Using the simplest case, the dependency for each of these factors are presented in this section.

3.4.1 Schottky Model of Metal-Semiconductor Contact

In the 1930s, Schottky developed the first acceptable theory on metal-semiconductor contacts [32]. In it, the rectifying behaviour observed in many contacts formed is attributed to the work function difference between metals and semiconductors [32]. (In a solid, the work function is the difference in energy between the fermi level and vacuum level.) Schottky's models for a metal-semiconductor barrier are shown in figure 3.14. The top part of figure 3.14 depicts the energy levels for metals and semiconductors

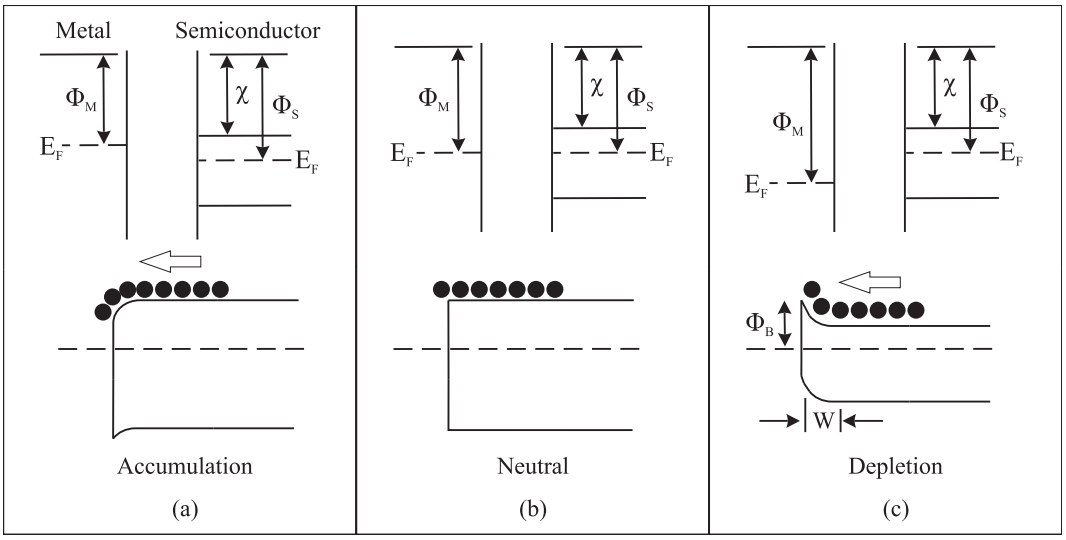


Figure 3.14: Energy levels formed between metal-semiconductor contacts based on Schottky's model [39]. Filled circles are electrons moving in the direction indicated by the arrow. No interfacial layers are present.

when not in contact, while the remaining half shows the energy levels when the two are in contact. Φ_M and Φ_S are the work function for metal and semiconductor respectively; χ the potential difference between the bottom of E_C and the vacuum level; and Φ_B the energy barrier seen by moving carriers (in an appropriate direction) defined as:

$$\Phi_B = \Phi_M - \chi \tag{3.42}$$

Depending on the flow direction of electrons and the E_F of respective materials, three contact properties can be formed. When both materials are in contact in (a), the

metal's E_F is lowered to the semiconductor's Fermi level. This "drags down" every other energy level in the metal in order for them to coincide with the corresponding semiconductor energy levels. If the movement of electrons is from the semiconductor to the metal, an *accumulation* type contact is formed, whereby electrons experience the least resistance in flowing out of the semiconductor [32]. A neutral contact is formed if the fermi level for both materials is the same (b). In this case, the contact is considered to be ohmic, as carriers travel unobstructed by energy barriers in both directions. Finally, if the metal's E_F is lower compared to the semiconductor's, a depletion contact is formed because electrons have to overcome an energy barrier of Φ_B to flow from the semiconductor to the metal (c).

When the modulator is in forward bias, electrons move through the following regions sequentially (assuming unity injection for both electrodes): cathode → 'n' doped Ge layer → intrinsic region → 'p' doped layer → anode. Hence a positive Φ_B is desired for a contact formed between the electrode and the 'n' doped surface (figure 3.14(c) but with electrons flowing in the *opposite* direction). The reverse is desired for contact between the 'p' doped layer and the anode since electrons move from the semiconductor to the metal.

3.4.2 Surface States

Dangling bonds and the species of chemicals adsorbed on the surface of Ge give rise to surface states. These states introduce energy levels that are within the energy gap of the host crystal, creating acceptor or donor type surface traps that could lead to an increase in Φ_B. To illustrate the detrimental effect of surface states, the energy levels for a contact formed between a metal-interfacial layer-n-type semiconductor with surface states (figure 3.15) is used [39]. Within Φ_0, states are of type donor, positively charged when empty and neutral when filled with electrons; while states above are of type acceptor, neutral when empty and negatively charged when filled with electrons. When acceptor states are filled with electrons, the resulting net surface trap charge density can be written as:

$$Q_{ss} = -qD_{it}(E_g - \Phi_0 - \Phi_B) \qquad (C/cm^2) \tag{3.43}$$

where D_{it} is the interfacial-trap density, and Φ_B the effective energy barrier formed between a metal and a semiconductor with an interfacial layer. Since these traps are charged (negatively in this case), a depletion layer similar to that in PN junction is

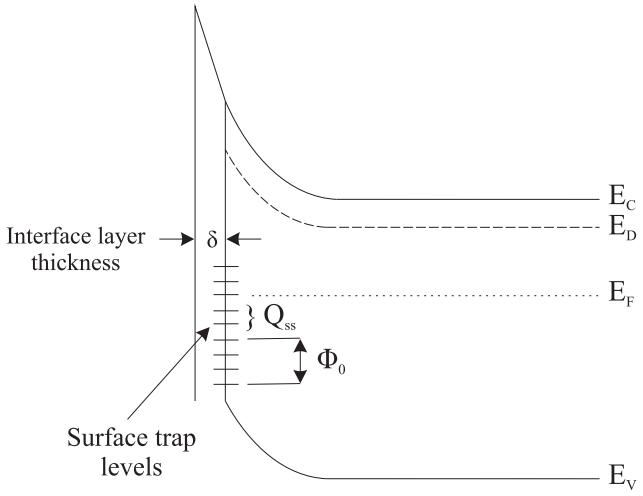


Figure 3.15: Energy diagram for contact between metal-interfacial layer-n-type semiconductor sandwich. The interfacial layer introduces states at the surface of the semiconductor.

formed in the bulk semiconductor. Inside this depletion layer, a positive space charge approximately equal to qN_d (N_d being the donor concentration in bulk semiconductor) creates a strong electric field at the surface. This field behaves similarly to Φ_B, which acts to oppose the flow of electrons directions where applicable.

Due to the chemically reactive nature of Ge surface (e. g., easily reacts with oxygen, forming oxides that dissolves even in water [42]), surface states formed are highly dependent on the type of chemicals the surface is exposed to. For example, the potential barrier between a metal-germanium contact could change by up to 0.5 V if the surface of germanium was first exposed to ozone prior to contact fabrication (surface becomes p-type) [17]; or it could be rendered n-type when treated with $H_2O + N_2$. Even exposure to ambient for several hours could shift Φ_B by 0.2 V. It must therefore be noted that information presented in this section are highly simplified, and only aims to highlight the adverse effects of surface states on the properties of metal-semiconductor contacts. (In fact, the electronic properties Ge surfaces have not been clearly understood. Even information on the chemical nature of Ge surfaces exposed to different chemicals is meagre [32, 43]).

3.4.3 Barrier Width Narrowing

From section 3.4.1, it is clear that an accumulation type contact is preferred, as electrons encounter the least barrier when moving in and out of the semiconductor. Un-

fortunately, this type of contact is difficult to form in real devices due to a phenomena called Fermi level pinning [32]. If the Fermi level of a semiconductor coincides with the top of Φ_0 (figure 3.15), the net surface interface charge becomes zero [39]. This energy level is likely to pin semiconductor E_F to a certain energy level. As a consequence, only small changes to E_F take place when semiconductors are in contact with metals. Barrier height engineering is therefore not a feasible method to produce low resistance contacts unless the the density of surface states can be controlled effectively—a feat not easily achieved even with modern day technology [31].

In practice, the number of electrons passing through a depletion type metal-semiconductor contact can be increased by reducing the depletion width W (figure 3.14). This increases the probability of electrons tunnelling through the metal-semiconductor energy barrier. Depletion width narrowing is achieved by increasing the doping concentration on the semiconductor, with a rough criteria for carrier tunnelling probability set by the characteristic energy E_{00} written as [39]:

$$E_{00} = \frac{qh}{4\pi}\sqrt{\frac{N\,(\mathrm{cm}^{-3})}{\varepsilon_s m^*_{tun}}} \qquad (\mathrm{eV}) \tag{3.44}$$

m^*_{tun} is the tunnelling effective mass for electron, ε_s the semiconductor permittivity, and N the density of dopants. For $E_{00} \approx kT$, electrons have to scale part of the energy barrier before tunnelling can take place. If $E_{00} > kT$, W is sufficiently narrow that electrons starts to tunnel near the bottom of E_C (figure 3.14) A high doping

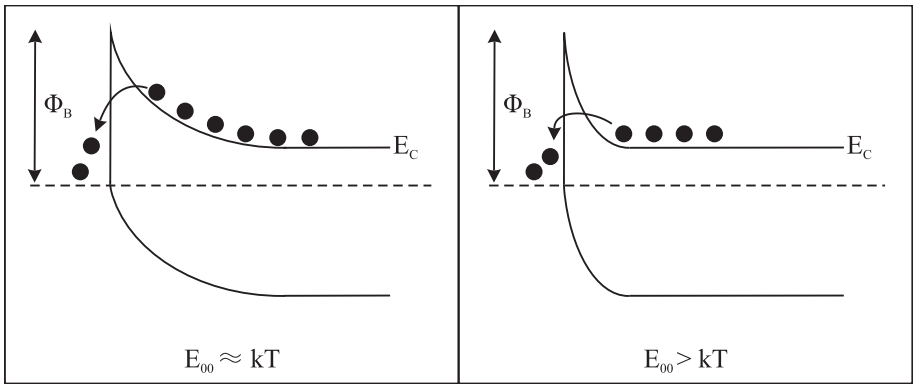


Figure 3.16: Carrier tunnelling criterion set by E_{00}. For $E_{00} > kT$, barrier becomes sufficiently narrow for tunnelling to take place close to the bottom of E_C

concentration at the surface is therefore desirable for producing low resistance contacts. Furthermore, it also augments the electrical performance of the modulator, as described in section 4.2.1.

3.4.4 Resistive Interfacial Layers

A resistive interfacial layer between metal and semiconductor is another factor that contributes to the increase in resistance in contacts formed. Native oxides, organic contaminants and particles (dust) on semiconductor surface all contribute to the interfacial layer. In the author's experience, a badly conditioned surface will significantly add to the total contact resistance. For example, the specific contact resistivity (ρ_c) for Al contacts on Ge could *easily* increase from 10^{-5} to $10^{-2}\,\Omega \cdot \mathrm{cm}^2$ if a thick layer of native oxide is present. This increase is considerable, since the total contact area is only $\approx 10^{-3}\,\mathrm{cm}^{-2}$ (an increase from $0.01\,\Omega$ to $10\,\Omega$ for a single contact). With a desired operating power $\approx 2\,\mathrm{W}$, the latter would limit current to less than $0.25\,\mathrm{A}$—an order of magnitude lower compared to the current required by the modulator to attenuate 95% of IR radiation at $10.59\,\mu\mathrm{m}$!

3.4.5 Discussion

Two methods can be deployed to produce low resistance contacts: barrier width narrowing using highly doped surfaces, and the removal of resistive interfacial layers that may introduce acceptor or donor type states at the surface. The latter is achieved by sequentially cleaning wafers with solvents and acids to remove organic and metallic contaminants, and also the ubiquitous layer of native oxide on Ge surface. Nevertheless, a very thin layer of oxide a few nanometers thick will always be present due to surface interaction with the ambient [44]. (Although the elimination of this layer might be possible using very advanced processing techniques, these were beyond the resources of this project.)

With respect to electrical contacts, the scope of our project only requires that they be made as ohmic as possible using standard cleanroom processing. Therefore, the approach taken by us is to combine the use of metal that: (a) theoretically forms low Φ_B with Ge, and (b) dopes Ge when its atoms diffuse into the crystal lattice at elevated temperatures (barrier width narrowing); and by using chemicals to remove interfacial layers from Ge surfaces prior metallization. No effort were taken to quantitatively study the aforementioned factors that contribute to electrical resistance *separately* (section 3.4.1 to 3.4.4). Rather, resistance for each factor are lumped together into a single term R_c, and the quality of contacts formed are evaluated based on that parameter. This is further elaborated in section 6.10.

A similar approach is adopted for device simulation using ATLAS: instead of imple-

menting complex models for energy height modifications due to work function difference and surface traps, fixed contact resistances from experimentally derived values of R_c are assigned for the anode and the cathode respectively. This approach is time saving and allows rapid assessment of the quality of metal-semiconductor contacts using the transmission line method (TLM). Section 6.10 details how easily R_C can be extracted using this technique. (Although TLM is a convenient method for characterizing contacts, structures on the surface have to be fabricated beforehand for TLM to be usable. This increases the overall time needed for analysis.)

Bibliography

[1] R. B. Adler, A. C. Smith, and R. L. Longini. *Introduction to Semiconductor Physics*, volume 1. John Wiley & Sons, New York, 1964.

[2] K. F. Berggren and B. E. Sernelius. Band-gap narrowing in heavily doped many-valley semiconductors. *Physical Review B*, 24(4):1971–1986, 1981.

[3] D. M. Caughey and R. E. Thomas. Carrier mobilities in silicon empirically related to doping and field. *Proceedings of the IEEE*, 55:2192–2193, 1967.

[4] P. P. Debye and E. M. Conwell. Electrical properties of n-type germanium. *Physical Review*, 93(4):693–706, 1954.

[5] W. C. Dunlap. *An Introduction to Semiconductors.* John Wiley & Sons, 1^{st} edition, 1960.

[6] C. Erginsoy. Neutral impurity scattering in semiconductors. *Physical Review*, 79:1013–1014, 1950.

[7] V. I. Fistul, M. I. Iglitsyn, and E. M. Amel'yanovskii. Mobility of electrons in germanium strongly doped with arsenic. *Soviet Physics Solid State*, 4(4):784–785, 1962.

[8] H. Fröhlich, B. V. Paranjape, C. G. Kuper, and S. Nakajima. The influence of interelectronic collisions and of surfaces on electronic conductivity. *Proceedings of the Physical Society*, 69:842–845, 1956.

[9] S. K. Ghandi. *Semiconductor Power Devices.* John Wiley & Sons, 1977.

[10] O. A. Golikova, B. Y. Moizhes, and L. S. Stil'bans. Hole mobility of germanium as a function of concentration and temperature. *Soviet Physics—Solid State*, 3(10):2259–2265, 1962.

[11] W. Gosling. Electronics. *Physics Education*, 14:400–406, 1979.

[12] A. S. Grove. *Physics and Technology of Semiconductor Devices.* Wiley, New York, 1967.

[13] J. B. Gunn. *Progress in Semiconductors*, 2:213, 1957.

[14] C. Jacaboni, F. Nava, C. Canali, and G. Ottaviani. Electron drift velocity and diffusivity in germanium. *Physical Review B*, 24(2):1014–1026, 1981.

[15] C. L. Julian and Jr. F. O. Lane. *Physical Review B*, 7(2):723–728, 1973.

[16] D. K. Kerry. Energy-gap narrowing and state filling in semiconductors under intense laser irradiation. *Physical Review B*, 18(12):7033–7037, 1978.

[17] R. H. Kingston. Review of germanium surface phenomena. *Journal of Applied Physics*, 27(2):101–114, 1956.

[18] G. G. MacFarlane, T. P. McLean, J. E. Quarrington, and V. Roberts. Fine structure in the absorption-edge spectrum of ge. *Physical Review*, 108(6):1377–1383, 1957.

[19] A. J. MacLeod. Algorithm 779: Fermi-dirac functions of order $-1/2, 1/2, 3/2, 5/2$. *ACM Transactions on Mathematical Software*, 24(1):1–12, 1998.

[20] J. L. Moll. *Physics of Semiconductors.* McGraw-Hill, 1^{st} edition, 1964.

[21] F. J. Morin and J. P. Maita. Conductivity and hall effect in the intrinsic range of germanium. *Physical Review*, 94(6), 1954.

[22] I. M. Nesmelova, N. I. Astaf'ev, and E. A. Nesmelov. Resistivity dependence of the absorption coefficient of crystalline germanium in the ir region. *Optical Materail Science and Technology*, 74(1):71–74, 2007.

[23] K. P. O'Donnell and X. Chen. Temperature dependence of semiconductor band gaps. *Applied Physics Letters*, 58(25):2924–2926, 1991.

[24] G. Ottaviani, C. Canali, F. Nava, and J. W. Mayer. Hole drift velocity in high purity ge between 8 and 220 kelvin. *Journal of Applied Physics*, 44(6):2917–2918, 1973.

[25] J. I. Pankove. *Optical Processes in Semiconductors.* Dover Publications, New Jersey, 2^{nd} edition, 1975.

[26] G. L. Pearson and J. Bardeen. Electrical properties of pure silicon and silicon alloys containing boron and phosphorus. *Physical Review*, 75(5):865–883, 1949.

[27] R. Pässler. Parameter sets due to fittings of the temperature dependencies of fundamental bandgaps in semiconductors. *Physica Status Solidi*, 216:975–1007, 1999.

[28] L. Reggiani, C. Canali, F. Nava, and G. Ottaviani. Hole drift velocity in germanium. *Physical Review B*, 16(6):2781–2791, 1977.

[29] I. M. Ross. The invention of the transistor. *Proceedings of the IEEE*, 86(1):7–28, 1998.

[30] A. Satta, E. Simoen, T. Clarysse, T. Janssens, A. Benedetti, B. D. Jaeger, M. Meuris, and W. Vandervorst. Diffusion, activation, and recrystallization of boron implanted in preamorphized and crystalline germanium. *Applied Physics Letters*, 87(17):172109, 2005.

[31] D. K. Schroder. The concept of generation and recombination lifetimes in semiconductors. *IEEE Transactions on Electron Devices*, 29:1336–1338, 1982.

[32] D. K. Schroder. *Semiconductor Material and Device Characterization.* John Wiley & Sons, New Jersey, 3^{rd} edition, 2006.

[33] K. Seeger. *Semiconductor Physics: An Introduction.* Springer-Verlag, Berlin, 6^{th} edition, 1997.

[34] B. E. Sernelius. Temperature-dependent resistivity of heavily doped silicon and germanium. *Physical Review B*, 41(5):3060–3068, 1990.

[35] Silvaco International, Santa Clara, CA, USA. *ATLAS User's Manual: Device Simulator Software*, 2004.

[36] R. A. Smith. *Semiconductors.* Cambridge University Press, Cambridge, 2^{nd} edition, 1978.

[37] S. M. Sze. *Physics of Semiconductor Devices.* John Wiley & Sons, 2nd edition, 1981.

[38] S. M. Sze and J. C. Irvin. Resistivity, mobility and impurity levels in gaas, ge and si at 300 k. *Solid-State Electronics*, 11:599–602, 1968.

[39] S. M. Sze and K. K. Ng. *Physics of Semiconductor Devices.* John Wiley & Sons, New Jersey, 3rd edition, 2007.

[40] C. D. Thurmond. The standard thermodynamic functions for the formation of electrons and holes in ge, si, gaas and gap. *Journal of The Electrochemical Society: Solid-State and Science Technology*, 122(8):1133–1141, 1975.

[41] Y. P. Varshni. *Physica*, 39:149, 1967.

[42] K. R. Williams, K. Gupta, and M. Wasilik. Etch rates for micromachining processing—part ii. *Journal of Microelectromechanical Systems*, 12(6):761–778, 2003.

[43] E. Yablonovitch, D. L. Allara, C. C. Chang, T. Gmitter, and T. B. Bright. Unusually low surface recombination velocity on silicon and germanium surfaces. *Physical Review Letters*, 57(2):249–252, 1986.

[44] P. V. Zant. *Microchip Fabrication.* McGraw Hill, 5th edition, 2004.

Chapter 4

The Influence of Carrier Lifetime on the Performance of the Modulator

The electrical characteristics of the modulator are a strong function of carrier lifetime. Its operating frequency (switching speed), achievable depth of modulation, power consumption and overall size (Ge thickness and aperture diameter) are parameters that are, within the desired modulator operating specifications, limited by carrier lifetime. For example, a short lifetime will set the upper limit on the allowed aperture size. Exceeding this limit will result in inhomogeneous absorption across the aperture, since the majority of diffusing carriers would have recombined before reaching the middle of the modulator.

Carrier lifetime has been studied extensively by various workers using simple analytical equations [22]. Though complex treatments on the dynamics of carriers are available [10, 24], analytical formulae have been shown to adequately model the behaviour of carriers subjected to various generation and recombination mechanisms in different devices (e. g., diodes, solar cells, transistors) [22]. Even so, literature pertaining information on carrier lifetimes must be treated with caution as many authors do not provide sufficient reference to details of device structure or measurement conditions [4, 16, 20, 21]. This could be misleading as carrier lifetime can easily change by orders of magnitude under different stimulus (e. g., low density electrical injection or high power above gap optical irradiation). Therefore, the subject of carrier recombination is first reviewed before analysis on factors influencing the performance of the modulator are explored later in this chapter.

4.1 Carrier Lifetime

Carrier lifetime falls into two categories: *generation* and *recombination* lifetimes denoted as τ_g and τ_r respectively [22]. Depending on the intended function of the device, τ_g may be more critical compared to τ_r. Since τ_g is only vital whenever there is a paucity of carriers (e. g., depletion region in reversed biased diodes), the recombination lifetime is more relevant to the modulator. For completeness, the generation lifetime is briefly reviewed.

4.1.1 Generation Lifetime

The generation lifetime τ_g can be written as a sum of three different processes as follows [16]:

$$\frac{1}{\tau_g} = \frac{1}{\tau_{eh}} + \frac{1}{\tau_o} + \frac{1}{\tau_{ii}}, \tag{4.1}$$

where τ_{eh}, τ_o and τ_{ii} are respectively lifetimes for carriers generated thermally, optically and via impact ionization. The effective generation rate of carriers, which is the inverse of τ_g, is important for devices where their functionality utilizes space-charge or depletion region. CMOS transistors and random access memory (RAMs) are such examples. To illustrate the dependency of the aforemention devices on generation lifetime, we examine specifically the expression for τ_{eh} given as [3]

$$\tau_{eh} = \tau_{p0}\left[\exp\left(\frac{E_T - E_i}{kT}\right)\right] + \tau_{n0}\left[\exp\left(\frac{E_i - E_T}{kT}\right)\right]. \tag{4.2}$$

τ_{n0} and τ_{p0} are the electron and hole lifetimes respectively; E_T the trap energy level and E_i the intrinsic fermi energy level. As carriers are generated thermally within depletion region in devices (to attain thermal equilibrium), they form a large fraction of leakage current with magnitude proportional to the density of electrically active impurities. In a CMOS transistors, this occurs in one of its channels during off state (either n or p-channel MOSFET), resulting in unwanted power dissipation. Leakage current is also the cause for low holding times in RAM as charge stored in its capacitor (formed between a polysilicon electrode and an inversion layer) is discharged by it. Consequently, a longer generation lifetime will result in lower leakage currents. This is achieved by reducing the number of impurities, or by using dopants that forms large $|E_T - E_i|$ [22].

Where excess carriers are present (Δn), the inverse mechanism for each of the generation terms in Eq. 4.1 becomes dominant. They are known as Shockley-Reed-Hall

(SRH), radiative and Auger recombination respectively. The following sections will be focussed on recombination process present in the modulator, since excess carriers are required for optical modulation. In particular, terminologies used for describing lifetimes at various operating regime are stressed. (Because the analysis of multi level traps are exceedingly complex; τ_{r} was treated using single level trap energy models. Again, this is the usual practice within the microelectronics community as it gives adequate representation of the cumulative effects of different recombination mechanisms on carrier lifetime [10, 16, 22, 23, 25].)

4.1.2 Recombination Lifetime

For any identifiable process that takes place at thermal equilibrium, there exist an inverse process of equal magnitude [1]. As a result, processes like the generation and recombination of thermally produced carriers are on average the same [10]. This is known as the *principle of detailed balance* [1]. If this balance is disturbed, say, by irradiating Ge with a steady stream of above gap energy photons, excess carriers are generated in the system that results in a new equilibrium state being established. The concentration of electrons and holes under this new state can be written as:

$$n = \overline{n} + \Delta n \tag{4.3}$$

$$p = \overline{p} + \Delta p \tag{4.4}$$

$\overline{n}$ and $\overline{p}$ are the thermal equilibrium concentrations for electrons and holes, while Δn and Δp are their time-dependent excess concentrations. Upon the removal of the stimulus (constant irradiation of photons), the excited system will revert back to thermal equilibrium through various recombination mechanisms [24]. The time required for this to take place is know as the recombination lifetime, and for an n-type semiconductor, the recombination rate can can be expressed as a power series [10]:

$$-\frac{\partial \Delta p}{\partial t} = \gamma_1 \Delta p + \gamma_2 \Delta p^2 + \gamma_3 \Delta p^3 + \ldots \tag{4.5}$$

Each term tend to dominate over different regimes of carrier concentration, but factors such as the chemical nature of dopants and its density must be taken into account for proper recombination lifetime analysis.

The fundamental recombination associated with each term in Eq. 4.5 are depicted in Fig. 4.1. (Only the simplest model for each mechanism are shown.) Because of charge neutrality requirement, the ratio between the number electrons and holes injected into

intrinsic Ge are close to unity. Therefore, we will use Δn to denote the concentration of both species of excess carriers.

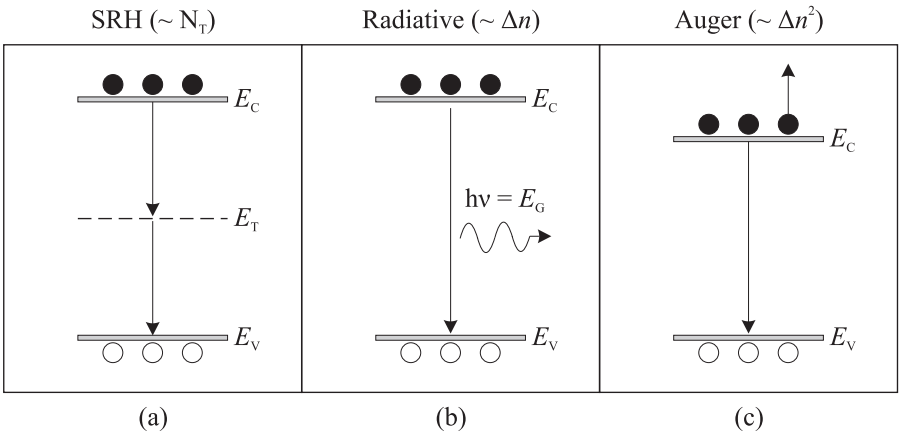


Figure 4.1: Fundamental recombination in semiconductors. (a) Deep level trap mediated carrier recombination (SRH). The energy from recombination is released as phonons. (b) Radiative recombination; (c) Auger recombination (filled circles represents electrons while unfilled circles, holes). The energy from recombination is transferred to another carrier, which may or may not require the interaction with phonons to conserve momentum. E_T, E_C, E_V are the trap energy level, the conduction band energy level and the valence band energy level respectively. SRH recombination is dependent on the density of traps (N_T), radiative recombination on (Δn), while Auger recombination on the square of excess carriers (Δn^2).

For a first order analysis, the effective recombination lifetime τ_{eff} has been defined as the time required for excess carriers to fall to $\frac{1}{e}$ times their original value [12]; and is evaluated by fitting the decay profile of carriers using the following equation:

$$\Delta n(t) = \Delta n(0) \exp\left(\frac{-t}{\tau_{\text{eff}}}\right) \tag{4.6}$$

where $\Delta n(0)$ is the initial value of excess carriers present. However, the assumption in Eq. 4.6 is not valid in systems where Auger recombination is considerable (Fig. 4.6). For intrinsic Ge, this occurs for concentrations larger than $10^{16}\,\text{cm}^{-3}$. A better way of representing recombination lifetime is to define it as follows [18]:

$$\tau_{\text{eff}} = \frac{\Delta n}{\Delta R} = \frac{\Delta n}{-\frac{\mathrm{d}\Delta n}{\mathrm{d}t}}, \tag{4.7}$$

where ΔR is the recombination rate of excess carriers and τ_{eff} the effective recombination lifetime that takes into account bulk and surface carrier lifetime

$$\frac{1}{\tau_{\text{eff}}} = \frac{1}{\tau_{\text{bulk}}} + \frac{1}{\tau_{\text{s}}}. \tag{4.8}$$

In this way, the instantaneous lifetime of carriers can be defined as a function of carrier concentrations. This is useful because only lifetimes for carrier concentrations in excess of $10^{16}\,\mathrm{cm}^{-3}$ are of interest to us. Recombination mechanisms responsible for the overall value of ΔR are presented next.

4.1.3 Shockley-Reed-Hall Recombination

SRH recombination describes the recombination of electrons and holes via intermediate impurity trap levels, with recombination energy released in the form of phonons (Fig. 4.1). The recombination lifetime is given as [22]:

$$\tau_{\mathrm{srh}} = \frac{\tau_{\mathrm{p}}(\Delta n + n_0 + n_0 \mathrm{e}^{E_{\mathrm{i}} - E_{\mathrm{T}}/kT}) + \tau_{\mathrm{n}}(\Delta n + p_0 + n_0 \mathrm{e}^{E_{\mathrm{i}} - E_{\mathrm{T}}/kT})}{p_0 + n_0 + \Delta n}. \tag{4.9}$$

n_0 and p_0 are the equilibrium electron and hole concentration respectively. τ_{p} and τ_{n} are given as:

$$\tau_{\mathrm{p}} = \frac{1}{\sigma_{\mathrm{p}}\,\nu_{\mathrm{th}}\,N_{\mathrm{T}}} \tag{4.10}$$

$$\tau_{\mathrm{n}} = \frac{1}{\sigma_{\mathrm{n}}\,\nu_{\mathrm{th}}\,N_{\mathrm{T}}}, \tag{4.11}$$

where σ_{p} and σ_{n}, N_{T} and ν_{th} are the cross sections for hole and electron capture by traps, the density of traps, and the thermal velocity ($\frac{1}{2}m^*\nu_{\mathrm{th}}^2 = 3/2kT$) respectively. For simplification, we define the following parameters:

$$A_{\mathrm{n}} = \sigma_{\mathrm{n}} \cdot \nu_{\mathrm{thn}} \tag{4.12}$$

$$A_{\mathrm{p}} = \sigma_{\mathrm{p}} \cdot \nu_{\mathrm{thp}} \tag{4.13}$$

Traps are formed by electrically active foreign elements inside semiconductors that introduce energy levels within the bandgap. Both dopants and impurities therefore contribute SRH recombination, with the former added intentionally while the latter, not. Equation 4.9 shows that SRH recombination is most effective for elements with trap energies close to E_{i}. It is for this reason why transition metals are so detrimental towards carrier lifetime: they introduce not only single, but multiple traps levels close to E_{i} when they occupy lattice sites substitutionally in Ge [23]. These traps are therefore known as deep level traps, while impurities with E_{T} close to E_{C} and E_{V} are known as shallow traps.

The capture rate for different energy levels introduced by the same TM varies. In single trap SRH recombination model, only a level with an activation energy closest to E_{i} is considered, since they are more efficient at aiding recombination (Eq. 4.9).

Table 4.2 lists several known TMs that are highly effective in reducing carrier lifetime in Ge. Differences in the chemical nature of each element are the reasons for the dissimilar capture rates listed.

Elements	E_{rel} (eV)	A_{n} ($\mathrm{cm^3\,s^{-1}}$)	A_{p} ($\mathrm{cm^3\,s^{-1}}$)
Cu	-0.26	4×10^{-10}	3.3×10^{-9}
	$+0.32$	3×10^{-10}	2×10^{-9}
Mn	-	$3{\times}10^{-9}\,\dagger$	$2 \times 10^{-9}\,\ddagger$
Co	-0.31	3×10^{-9}	-
	$+0.25$	3×10^{-8}	-
Fe	-0.27	3×10^{-9}	2×10^{-7}
	$+0.34$	3×10^{-8}	6×10^{-8}
Ni	-0.30	$6.5 \times 10^{-8}\,*$	3.9×10^{-7}
	$+0.22$	1.5×10^{-9}	

Figure 4.2: The capture rate for electrons in p-type Ge (A_{n}) and for holes in n-type Ge (A_{p}). Negative values for E_{rel} are defined relative to E_{C} while positive values with respect to E_{V}. Data shown were compiled by Simoen *et. al.* from various sources [23]. ($*$) The high value of A_{p} deviates from the expected trend and requires more detailed analysis [23]. ($\dagger$,$\ddagger$) corresponds to different energy levels.

Although copper is a well known contaminant in Ge owing to its high diffusion rate and mid gap trap levels, recent work has shown that, at similar densities, Fe, Ni and Co are more efficient lifetime 'killers'[23]. Both A_{n} and A_{p} of Ni for example are at least two orders of magnitude larger compared to Cu. Nevertheless, because Cu is more prevalent in Ge due to its high diffusivity [25], its SRH related recombination parameters were used in ATLAS to model the electrical characteristics of the modulator with small impurity content.

It is evident from Eq. 4.9 that τ_{srh} varies for different carrier injection levels (i. e., different Δn). Assuming an intrinsic Ge sample with $10^{11}\,\mathrm{cm^{-3}}$ of Cu impurities, the dependence of τ_{srh} on Δn is plotted in Fig. 4.3. In the limit of high carrier injection, the lifetime can be expressed as [22]:

$$\tau_{\mathrm{hl}} = \tau_{\mathrm{n}} + \tau_{\mathrm{p}}, \tag{4.14}$$

and is independent of E_{T}. τ_{hl} is also known as the ambipolar lifetime, which refers to lifetime of carriers at injection levels much higher compared to the density of traps.

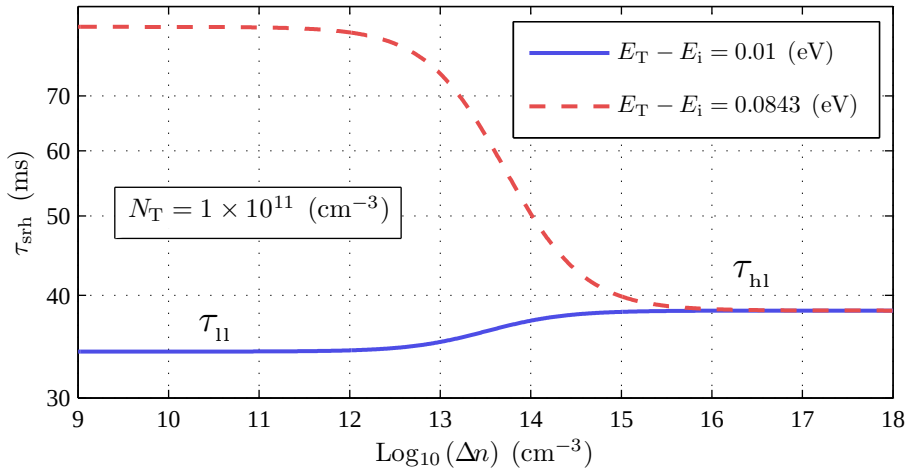


Figure 4.3: SRH recombination lifetime as a function injected carrier density. E_{rel} of 0.26 eV (dotted line) and 0.32 eV (solid line) were used.

4.1.4 Radiative Recombination

In radiative recombination, electron hole pairs recombine from band to band, and the resulting change in energy is released in the form of photons (Fig. 4.1) [26]. For Ge, both direct and indirect recombination occur simultaneously, each releasing photons with wavelength centred at 1.5 μm and 1.77 μm respectively. The radiative lifetime is expressed as [22]:

$$\tau_{rad} = \frac{1}{B(p_0 + n_0 + \Delta n)}, \tag{4.15}$$

where B is the radiative recombination probability with dimensions cm^3/s. This probability is a function of intrinsic carrier concentration

$$B = \frac{R}{n_i^2}. \tag{4.16}$$

R is the thermal equilibrium recombination rate derived from the van Roosbroek-Shockley theory on radiative recombination. (A lengthy review on the derivation of R at thermal equilibrium have been conducted by Varshni [27]; thus, only relevant conclusions from this theory are presented.)

Unlike Auger recombination, the value for R has not changed much ever since it was investigated in the early fifties [5, 22, 27]. Values between 1.57 and 2.5 (cm^{-3}s^{-1}) have been reported and this is in good agreement with van Roosbroek-Shockley's theory on radiative recombination [12, 27]. If excess carriers are present in non-degenerate Ge, the increase in R is given as [27]:

$$R' = \frac{\Delta n \Delta p}{n_i^2} R, \tag{4.17}$$

since the rate of recombination must be proportional to the number of carriers present. However, published values of R are only valid at thermal equilibrium (or for small deviations from it as given in Eq. 4.17), and for samples less than 150 μm thick[12, 27]. For high carrier densities in thick samples ($\alpha d > 1$), the evaluation of recombination rate becomes complicated due to photon recycling. In this process, photons generated from recombination are reabsorbed inside the lattice, effectively reducing the effect of recombination[2]. It has been suggested that radiative recombination is a factor that will fundamentally limit the maximum obtainable carrier lifetime for semiconductors at high injection levels, until long carrier lifetimes observed in materials such as GaAs were explained by photon recycling [7]. However, this argument does have its validity for Ge. As mentioned previously, two emission peaks have been observed in Ge: one centred at 1.5 μm and the other at 1.77 μm. Since Ge is an indirect gap semiconductor, its absorption coefficient rapidly reduces at several kT above the absorption edge ($\sim 10\,\mathrm{cm}^{-1}$) [7]. Photons therefore stand a reasonable chance of escaping the crystal since the FWHM of the indirect peak is around 260 nm.

In practice, it is customary to ignore the contribution of radiative recombination in Si and Ge as its effect is negligibly small compared to other recombination mechanisms at low Δn [10, 22]. But for devices with broad carrier injection range and long carrier lifetimes such as PiN diodes, Eq. 4.15 shows that below Δn of $10^{17}\,\mathrm{cm}^{-3}$, radiative recombination is considerable for Ge (Fig. 4.4). We have erroneously assumed the same until we decided to evaluate its contribution in our device. Frustratingly, many published works on carrier lifetimes in Ge—and even in standard high power device reference books—ignore the contribution from radiative recombination without giving proper justification for it [9, 10, 20]. This could be due to τ_{rad} being significant for only a narrow range of carrier concentration. Unfortunately, τ_{rad} is important for the performance of the modulator, as carrier densities ranging from $3 - 6 \times 10^{16}\,\mathrm{cm}^{-3}$ are needed to achieve $> 95\%$ IR attenuation. The calculated instantaneous carrier lifetime for various bulk recombination mechanisms is shown in Fig. 4.4.

A reasonable way of calculating τ_{rad} (in the author's opinion) is to consider only indirect recombination rates (R_{i}), assuming 100% of photons generated via direct transition are reabsorbed. Table 4.1 (a) and (b) list all relevant parameters for radiative recombination. But to be on the safe side, we have used R instead of R_{i} in simulation work . This approach is pessimistic (and overly so) but nevertheless, ATLAS has shown that modulator design can still be optimized to achieve beyond 95% optical

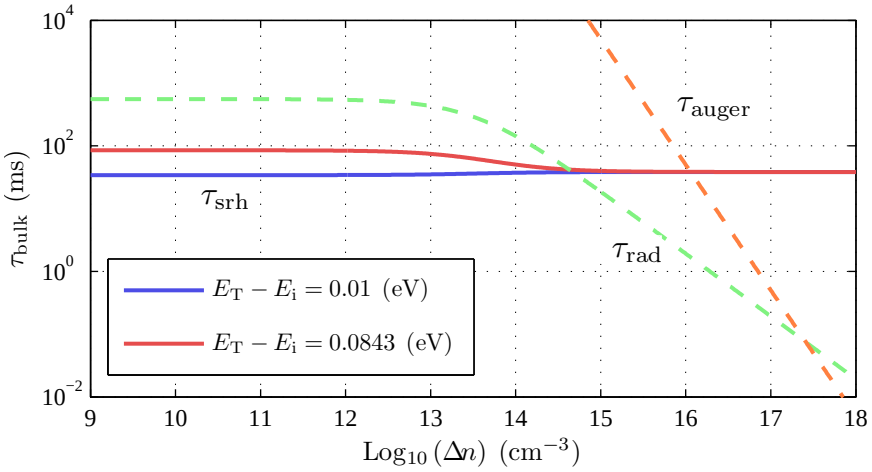


Figure 4.4: The instantaneous bulk recombination lifetimes for different mechanisms in Ge. B of $5.2\,\text{cm}^3/\text{s}$, and C_n and C_p of $10^{-31}\,\text{cm}^6/\text{s}$ have been used. (The exact range of concentration at which radiative recombination is dominant could not be define accurately. Please see discussions in sections 4.1.4 and 7.4.1 for more details.)

modulation. These are explored in Ch. 5.

4.1.5 Auger Recombination

Auger recombination is by far the most detrimental recombination mechanism acting on the modulator at high carrier densities ($> 10^{17}\,\text{cm}^{-3}$). It involves the transfer of energy to another carrier (either electron or hole) when an electron-hole pair recombine, creating highly energetic charged carrier[24]. In intrinsic material, Auger recombination is characterized by:

$$-\frac{\text{d}\Delta n}{\text{d}t} = (C_\text{n} + C_\text{p})\Delta n^2, \tag{4.18}$$

where C_n and C_p are the Auger coefficients for e-e-h (energy transferred to electron) and h-h-e (energy transferred to hole) processes respectively. Since both coefficients are very similar in Ge [22], they can be added to form γ, and the instantaneous carrier lifetime for Auger recombination written as:

$$\tau_\text{aug} = \frac{1}{\gamma \cdot \Delta n^2}. \tag{4.19}$$

It was not until recently that experimental values for γ converged to values between 1 and $4 \times 10^{-31}\,\text{cm}^6/\text{s}$ for non-degenerate Ge [6, 9, 19, 20, 22]. Previously, theoretical and experimental figures for γ could differ by two orders of magnitude (from 10^{-32} to $10^{-30}\,\text{cm}^6/\text{s}$) [6, 11, 15]. This posed considerable headache to us because if γ was $10^{-30}\,\text{cm}^6/\text{s}$, the performance of the modulator will be crippled unless very thick Ge

Element	T	R	n_{i}	B	τ
	(^oK)	$(\mathrm{cm^{-3}s^{-1}})$	$(\mathrm{cm^{-3}})$	$(\mathrm{cm^3\,s^{-1}})$	(s)
Ge	77	2.4×10^{-26}	2.42×10^{-7}	4.1×10^{-13}	5.04×10^{18}
	300	2.85×10^{13}	2.33×10^{13}	5.25×10^{-14}	0.409

(a)

Element		Indirect		Direct	
	T	R_{i}	B_{i}	R_{d}	B_{d}
	(^oK)	$(\mathrm{cm^{-3}s^{-1}})$	$(\mathrm{cm^{-3}})$	$(\mathrm{cm^3\,s^{-1}})$	(s)
Ge	77	1.9×10^{-27}	6.8×10^{-14}	1.5×10^{-32}	5.2×10^{-19}
	300	4.7×10^{12}	1.5×10^{-13}	1.9×10^{13}	6.4×10^{-13}

(b)

Table 4.1: Radiative recombination parameters by Varshni [27] calculated using van Roosbroeck-Shockley theory from data obtained experimentally

($> 2\,\mathrm{mm}$) is used (section 5.4.4). And unlike SRH recombination whose rate can be controlled by limiting the number of impurities through careful processing, Auger is an intrinsic property of the material and cannot be changed. Its square dependency on carrier concentration rapidly reduces lifetime at high densities, a condition true for regions beneath the electrodes where carriers easily exceed $10^{17}\,\mathrm{cm^{-3}}$. The carrier lifetimes for different Auger recombination are plotted in Fig. 4.5. As can be seen, lifetime drop exponentially as the square of Δn.

A γ of $10^{-31}\,\mathrm{cm^6/s}$ was used for simulating the electrical behaviour of the modulator. There is no reason to adopt pessimistic values for γ since a large number of work have concluded that it does not exceed $4 \times 10^{-31}\,\mathrm{cm^6/s}$ [6, 9, 19, 20, 22]. In fact, the majority of results converge to values close to $10^{-31}\,\mathrm{cm^6/s}$.

With all relevant bulk recombination defined, the effective bulk lifetime for carriers in Ge can be written as:

$$\frac{1}{\tau_{\mathrm{bulk}}} = \left(\frac{1}{\tau_{\mathrm{srh}}} + \frac{1}{\tau_{\mathrm{rad}}} + \frac{1}{\tau_{\mathrm{aug}}} \right). \tag{4.20}$$

τ_{bulk} is plotted against excess carriers in Fig. 4.6. As we have stressed at the beginning, carrier lifetime is highly dependent on Δn. This is crucial for analysing the performance of the modulator during 'on' state because to achieve good depth of modulation, Δn larger than $10^{16}\,\mathrm{cm^{-3}}$ are needed. Therefore, only lifetimes that corresponds to the actual carrier concentration matters. Nevertheless, the increase in carrier lifetime

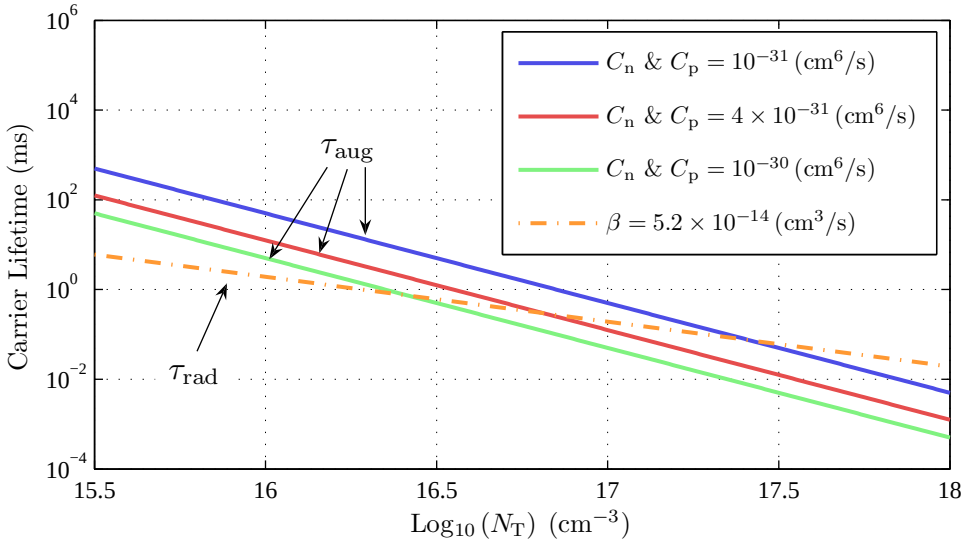


Figure 4.5: Carrier lifetime for various Auger coefficients

at low Δn is advantageous for conserving power needed to maintain specific level of modulation (i. e., carriers can be replenished more easily).

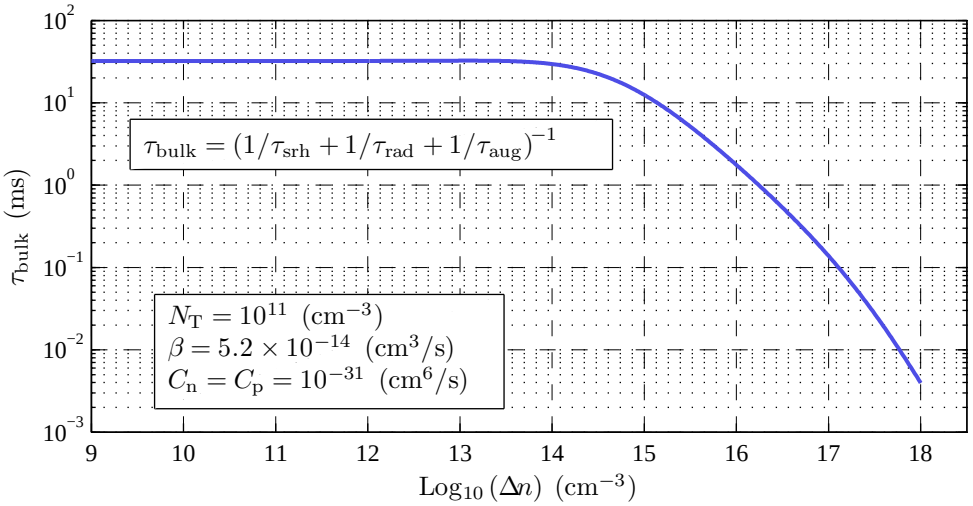


Figure 4.6: Bulk carrier lifetimes for intrinsic Ge. At high carrier concentrations, lifetime becomes non-exponential due to Auger recombination.

The final recombination mechanism relevant to the modulator is presented next.

4.1.6 Surface Recombination

In addition to bulk recombination, carriers can recombine via intermediate trap levels formed at the surface [22]. The density of these traps have been attributed to the formation of oxide, and to lattice damage from mechanical polishing [21]. However,

studies conducted by Fairley, Ponpon and Yablanovitch *et. al.*, to name a few, show that recombination at the surface depends strongly on the density of dangling bonds and also the species of chemicals present at the surface [8, 21, 28]. These form traps on the surface that, under certain conditions, cause surface recombination to exceed its bulk counterpart [24].

Surface recombination is quantified using the parameter 's', and is better known as the surface recombination velocity (SRV). It is defined as the ratio of surface recombination rate per unit area to Δn at the surface. For Ge, SRV ranging from 2 to $10,000\,\mathrm{cm/s}$ have been reported [8, 28]. For our interest, low recombination velocity is desired to preserve carrier lifetime. One way to conveniently achieve this is by etching the surface with suitable acids. More information on producing low SRV surfaces are presented in Chapter. 6. In it, the dependencies and time variation of SRV in Ge are also discussed.

Surface recombination have been modelled using the SRH recombination equation [28]. The recombination lifetime is given as [22]:

$$\tau_\mathrm{s} = \frac{s_\mathrm{n}(n_\mathrm{s} + n_\mathrm{i}\exp\left(\frac{E_\mathrm{T}-E_\mathrm{is}}{kT}\right) + s_\mathrm{p}(p_\mathrm{s} + n_\mathrm{i}\exp\left(\frac{-E_\mathrm{T}-E_\mathrm{is}}{kT}\right))}{s_\mathrm{n}s_\mathrm{p}(p_\mathrm{s}n_\mathrm{s} - n_\mathrm{i}^2)}, \tag{4.21}$$

where s_n and s_p are $\sigma_\mathrm{ns}\nu_\mathrm{th}N_\mathrm{it}$ and $\sigma_\mathrm{ps}\nu_\mathrm{th}N_\mathrm{it}$ respectively. Every parameter denoted with subscript 's' in Eq. 4.21 corresponds to ones used in SRH recombination model, but are applicable only to conditions at the surface . Readers are referred to Schroder [22] for more information on the use of Eq. 4.21 for evaluating surface carrier lifetime. For practical purposes, the formula that predicts the combined recombination effects of two surfaces for a specified wafer thickness is preferred. In this way, the effects of bulk and surface recombination on carrier lifetime can be evaluated. This was given by Grivickas *et. el.* and their formula has an uncertainty of 4% over broad range of 's' values [17]. It is written as:

$$\tau_\mathrm{s} = \frac{d}{2s} + \frac{d^2}{D_\mathrm{a}\pi^2}, \tag{4.22}$$

where D_a is the ambipolar diffusion coefficient and d is the wafer thickness. Each surface is assumed to have the same 's'. D_a is fairly constant for Δn between 10^{13} and $10^{18}\,\mathrm{cm}^{-3}$, but increases rapidly beyond $10^{19}\,\mathrm{cm}^{-3}$ [29]. It is expressed as [10]:

$$D_\mathrm{a} = \frac{2D_\mathrm{n}D_\mathrm{p}}{D_\mathrm{n} + D_\mathrm{p}}. \tag{4.23}$$

D_n and D_p have values of 103 and $54\,\mathrm{cm}^2/\mathrm{s}$ respectively. Subscript 'n' refers the coefficient for electrons while is 'p' for holes. Having defined τ_s, the effective carrier

lifetime can be evaluated as a function of wafer thickness using

$$\tau_{\text{eff}} = \left(\frac{1}{\tau_{\text{bulk}}} + \frac{1}{\tau_{\text{s}}} \right)^{-1}. \tag{4.24}$$

Using τ_{bulk} of 0.1, 0.3 and 0.5 ms, where each corresponds to lifetimes for Δn at 10^{17}, 3×10^{16} and $10^{16}\,\text{cm}^{-3}$ respectively, τ_{eff} for associated τ_{bulk} are plotted in Fig. 4.7.(As mentioned previously, only lifetimes at carrier densities present during modulator 'on' state matters, hence the choice of τ_{bulk}.) Using acid etch, stable SRV values from 100

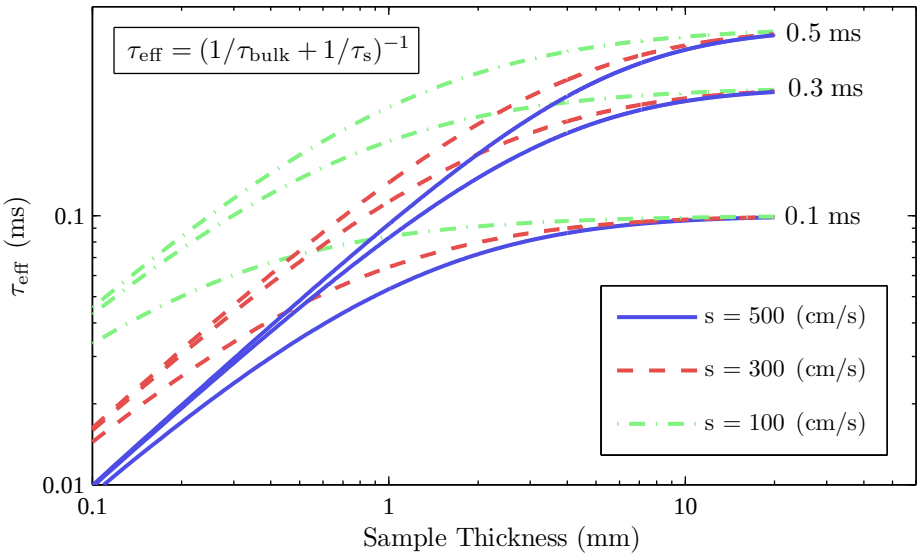


Figure 4.7: Effective carrier lifetime as a function of SRV and wafer thickness

to 300 cm/s can be obtained [21]. But even at these levels, the deleterious effect of surface recombination is still significant for thicknesses less than 2 mm. For very thick samples, τ_{eff} approaches τ_{bulk}.

4.2 Electrical Characteristics of PiN Diode

Parameters that affect the performance of PiN diode are explored in this section. For this purpose, simple analytical equations for a 1D diode structure are utilized. These equations provide insight into the electrical behaviour of diodes, important for aiding the design of optimized modulator structures in ATLAS (Ch. 5).

4.2.1 Forward Biased Diode

The modulator is essentially a big piece of Ge made into PiN diode. It is the preferred structure for electrically injecting carriers into Ge because the thick 'i' region acts as a

large reservoir for carriers needed for modulation. Because this region consist of intrinsic material, SRH recombination would be minimal. Furthermore, provided the integral of carriers through the modulator exceeds $5 \times 10^{15}\,\mathrm{cm}^{-2}$ (for $\approx 95\%$ modulation), it can be made thick to reduce the overall density of carriers in the crystal, mitigating the effect of Auger and radiative recombination. However, the maximum thickness allowed (within the desired operating specifications) is limited by the diffusion length of the carrier defined as:

$$L_\mathrm{a} = \sqrt{D_\mathrm{a}\tau_\mathrm{eff}}. \tag{4.25}$$

This length can be viewed as the distance carriers can diffuse before they are annihilated [25]. Assuming a 2 mm thick wafer, the diffusion length for different bulk lifetimes (each corresponding to specific carrier density shown in Fig. 4.6) are plotted in Fig. 4.8.

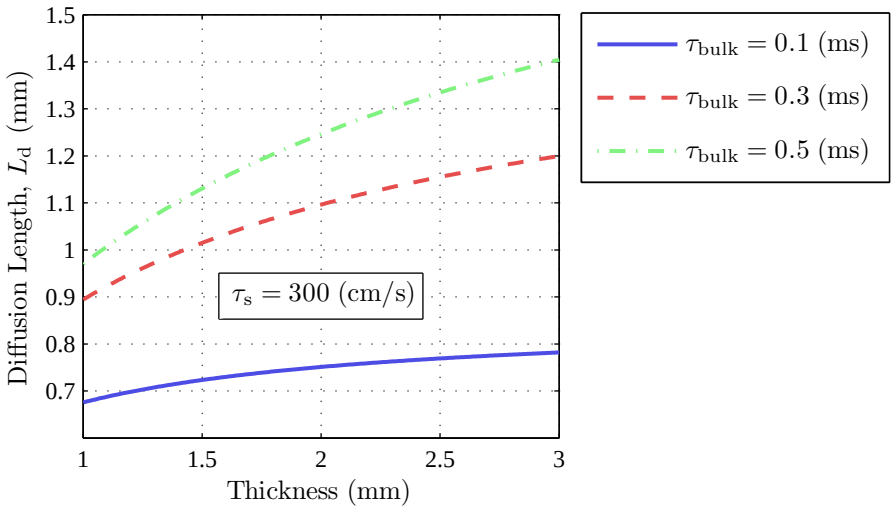


Figure 4.8: Carrier diffusion length as a function of τ_eff

To convey the usefulness of L_d, an abrupt junction with 'i' region the width of $2d$ (Fig. 4.9) is investigated under forward bias conditions. Two components of current

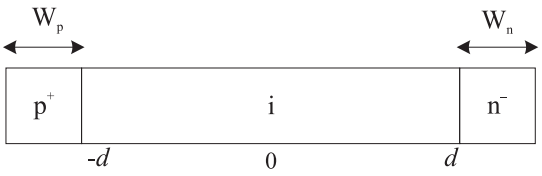


Figure 4.9: A PiN diode structure

density exist at forward bias: the drift component, which is the movement of charge driven by electric field, and the diffusion component caused by the gradient in carrier concentration[26]. (Because carriers are constrained to one direction in Fig. 4.9, side

way diffusion of carriers is not shown. A larger sample will be used to illustrate diffusion currents). In our one dimensional example, these are written as:

$$J_{\mathrm{n}} = q\mu_{\mathrm{n}}\Delta n\varepsilon + qD_{\mathrm{n}}\frac{\mathrm{d}(\Delta n)}{\mathrm{d}x} \tag{4.26}$$

$$J_{\mathrm{p}} = q\mu_{\mathrm{p}}\Delta p\varepsilon + qD_{\mathrm{p}}\frac{\mathrm{d}(\Delta p)}{\mathrm{d}x}, \tag{4.27}$$

where ε is the electric field (V/cm), μ the mobility of carriers ($\mathrm{cm}^2/\mathrm{V}\cdot\mathrm{s}$), and D the diffusion coefficient (cm^2/s). Subscripts 'n' and 'p' denote properties related to electrons and holes respectively. The first term in Eq. 4.26 and 4.27 represent the drift component while the second for the diffusion component. For carriers generated via electrical injection, the source of diffusing carriers come from the drift currents through the p^+-i-n^- regions at forward bias. Since the diffusion rate is dependent on the gradient of carrier concentration, and a Δn of more than $2\times10^{16}\,\mathrm{cm}^{-3}$ is needed for $>$ 95% modulation in a 2 mm thick wafer, the concentration of carriers in the drift component should be a few factors larger to drive diffusion. For a fixed current density (J_{a}), the concentration of carriers through the diode can be expressed using the formula [10]

$$\Delta n(x) = \frac{\tau_{\mathrm{eff}}J_{\mathrm{a}}}{2qL_{\mathrm{a}}}\left[\frac{\cosh(x/L_{\mathrm{a}})}{\sinh(d/L_{\mathrm{a}})} - B\frac{\sinh(x/L_{\mathrm{a}})}{\cosh(d/L_{\mathrm{a}})}\right], \tag{4.28}$$

where B is defined as:

$$B = \frac{\mu_{\mathrm{n}} - \mu_{\mathrm{p}}}{\mu_{\mathrm{n}} + \mu_{\mathrm{p}}}. \tag{4.29}$$

μ_{n} and μ_{p} are 3900 and $1900\,\mathrm{cm}^2/\mathrm{V}\cdot\mathrm{s}$ respectively. The profiles of Δn are plotted in Fig. 4.10(a) using Eq. 4.28. Notice how the difference in carrier mobility shifts the concentration minima towards the right via the parameter B. As a comparison, a 3D model of the modulator at similar thickness, driven with 2.6 A of forward current is plotted alongside it. Similarities between curves generated give confidence to the use of analytical equation in analyzing the electrical characteristics of a diode qualitatively. From Eq. 4.28, a large τ_{eff} leads to higher carrier concentrations along the diode, while long L_{a} produces a flatter, more homogeneous distribution of carriers. Too short an L_{a} will result in severe reduction in diode conductivity modulation (Fig. 4.11), as ohmic resistance rapidly increases in the 'i' region.

For a given forward voltage, (V_{F}), the total potential across the diode can be written as a sum of potential drops in three regions:

$$V_{\mathrm{T}} = V_{\mathrm{L}} + V_{\mathrm{m}} + V_{\mathrm{R}}, \tag{4.30}$$

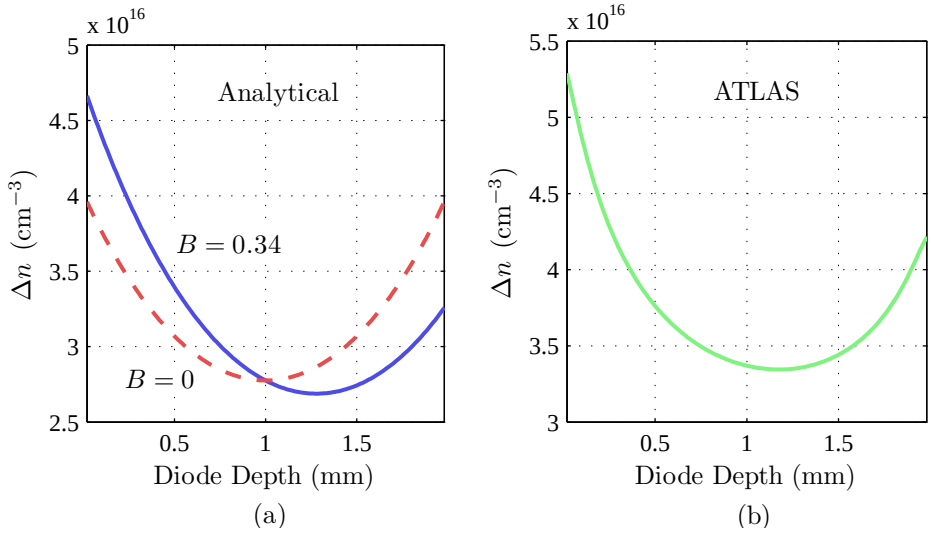


Figure 4.10: (a) Carrier concentration through a 2 mm thick diode at $J_A = 6\,\mathrm{cm}^2/\mathrm{A}$, calculated using Eq. 4.28; (b) Δn for a 3D model of the modulator driven with 2.6 A

Where V_L and V_R are voltage drops around p^+ and n^- regions respectively. The potential drop across the 'i' region (V_m) is estimated using [10]

$$V_m = \left\{ \frac{8b}{(b+1)^2} \frac{\sinh(d/L_a)}{\sqrt{1 - B^2 \tanh^2(d/L_a)}} \cdot \arctan\left[\sqrt{1 - B^2 \tanh^2\left(\frac{d}{L_a}\right)} \sinh\left(\frac{d}{L_a}\right) \right] \right\} + B \ln\left[\frac{1 + B \tanh^2(d/L_a)}{1 - B \tanh^2(d/L_a)} \right], \tag{4.31}$$

where $b = \mu_n/\mu_p$. It is interesting to note that the voltage drop across the 'i' region is independent of current through it. It does however show that long carrier lifetimes are crucial to keep power consumption low during forward bias. Plotting V_m against d/L_a (Fig. 4.11), one could see that if wafers are thicker than L_a, or if L_a is very short due to low τ_{eff}, a steep rise in voltage drop across the intrinsic region will result. As mentioned previously, a thick wafer can be used to reduce the effects of Auger and radiative recombination. This however works only for, assuming static τ_{bulk}, $d \leq 2L_a$. However, simulations using ATLAS shows that for our device, improvements are still obtained even for a 4 mm thick wafer, where $d/L_a > 2$ (Sec 5.4.4). This is due to our assumption that τ_{bulk} remains static, whereas in real devices, it increases as the concentration of carrier drops (Fig. 4.6).

With J_a and V_m defined, and if an upper limit for total power dissipation is set,

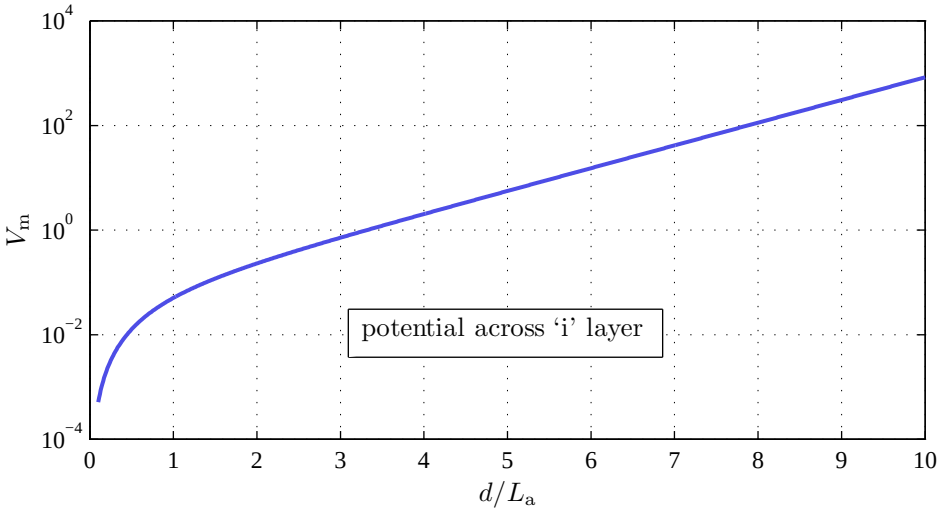


Figure 4.11: Potential drop across the 'i' region as a function of d/L_a.

say P_{max}, one could estimate the area of electrode required using:

$$P_{max} = V_m J_a A, \tag{4.32}$$

where A is the area for electrical contact. Unfortunately, this assumption is incorrect for real devices as current densities around the p^+ and n^- region can be very high [10]. This is illustrated in Fig. 4.12. $\Delta n(x)$ plotted in dotted line in Fig. 4.10 uses constant J_a at 6 A/cm^2. To produce similar density of carriers in the mid section of the modulator, higher current densities are needed—as result from ATLAS clearly shows. Close to

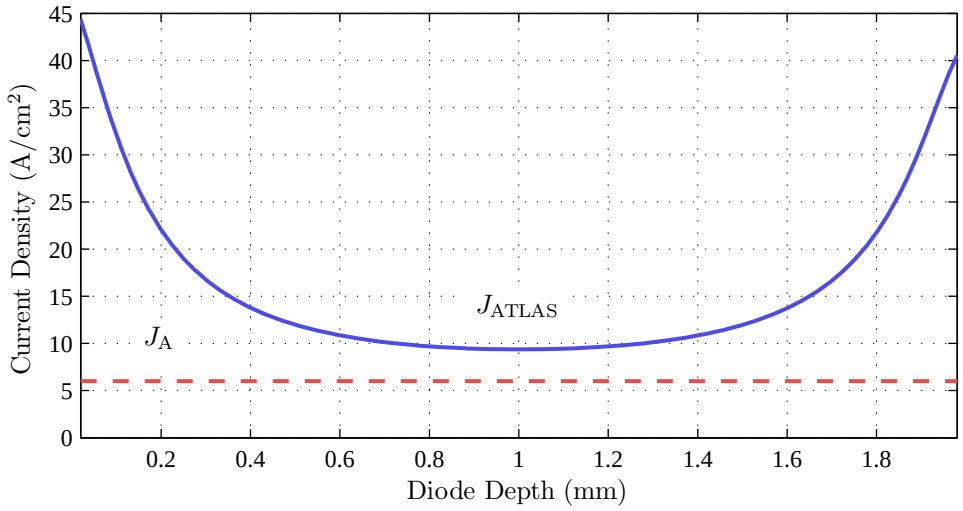


Figure 4.12: Current density through PiN diodes estimated using: (a)analytical equations and (b)using ATLAS

p^+ and n^-, J_{ATLAS} started out really high but decays rapidly thereafter. This is due to Auger recombination, and to minority carriers being injected into the p^+ (electron

injection) and n^- (hole injection) region. This minority injection does not contribute to conductivity modulation (i. e., does not add to the number of carriers in the diode during forward bias), but greatly increases the voltage drop across the diode[13, 14]. The current density for minority carrier injection are given as:

$$J_{\mathrm{nP}} = \left[\frac{qD_{\mathrm{nP}}n_{\mathrm{p}}}{L_{\mathrm{nP}}\tanh(W_{\mathrm{p}}/L_{\mathrm{nP}})}\right] \mathrm{e}^{qV_{\mathrm{L}}/kT} \tag{4.33}$$

$$J_{\mathrm{pN}} = \left[\frac{qD_{\mathrm{pN}}p_{\mathrm{n}}}{L_{\mathrm{pN}}\tanh(W_{\mathrm{n}}/L_{\mathrm{pN}})}\right] \mathrm{e}^{qV_{\mathrm{R}}/kT} \tag{4.34}$$

where n_{p}, D_{nP}, L_{nP} and W_{p} are parameters for the p^+ region and p_{n}, D_{pN}, L_{pN} and W_{n} for n^- region. W is the depletion width formed between the intrinsic region and the respective doped region.

Because of minority injection, carrier injection efficiency is degraded; thus more potential must be applied during forward bias to reach desired current modulation in the 'i' region. Furthermore, carriers injected into the 'i' region are subjected to Auger recombination, further lowering Δn. One way to reduce minority injection is to increase the concentration of holes in the p^+ region. This reduces the number of electrons n_{p}, since charge neutrality requires that $np = n_{\mathrm{i}}^2$, which in turn lowers J_{nP}. The same applies to J_{pN} except that p_{n} must be reduced instead. Therefore, highly doped p^+ and n^- region are crucial for low power dissipation in the modulator. In Sec. 5.4.2, the current and voltage characteristic of the modulator as a function of doping density are given, thus we will not elaborate on this matter further.

Having established parameters influencing drift currents, we now look at the resulting diffusion component and its dependence on carrier lifetime. For that, the structure shown in Fig. 4.13 is used. Assuming a fixed Δn for the drift component (red arrow),

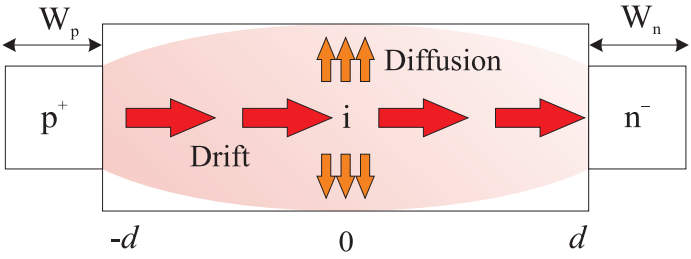


Figure 4.13: Drift and diffusion current components in a PiN diode

the concentration of carriers that constitute the diffusion currents in either the up or downward direction (brown arrow) can be evaluated using:

$$\Delta n_{\mathrm{diff}}(x) = n_{\mathrm{i}} + [\Delta n - n_{\mathrm{i}}] \exp\left(-\frac{x}{L_{\mathrm{a}}}\right), \tag{4.35}$$

A long L_a enables larger number of carriers to diffuse further. If $\Delta n = 4 \times 10^{16}\,\mathrm{cm}^{-3}$, the profiles for wafers with different thickness are shown in Fig. 4.14.

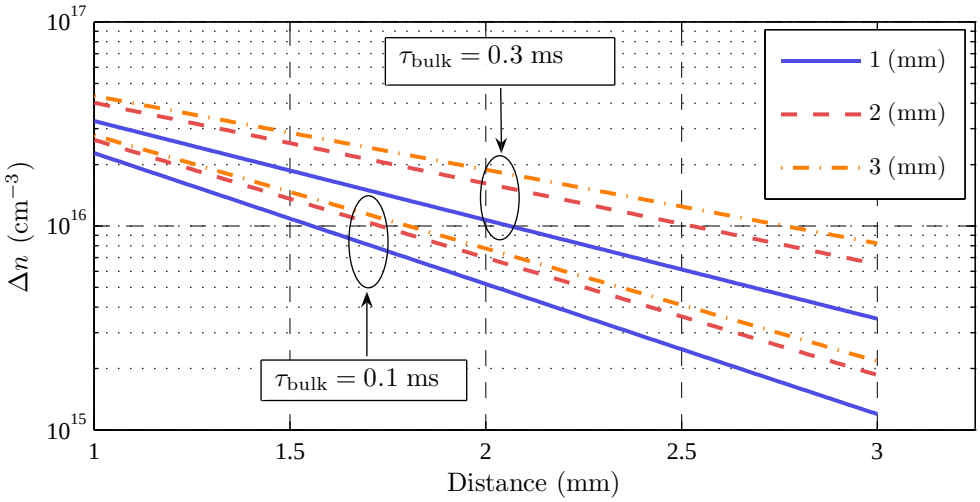


Figure 4.14: Carrier concentration as a function of distance away from a fixed source of $\Delta n = 4 \times 10^{16}\,\mathrm{cm}^{-3}$. SRV was set to 300 cm/s.

Marked improvement is obtained if the thickness of the wafer is increased from 1 to 2 mm. Beyond 2 mm, the severity of recombination mechanisms are reduced, leading to ever smaller improvements.

4.2.2 Discussion

We have highlighted the importance of specifying bulk carrier lifetime in accordance with the density of carriers present; for the reason that it may change by orders of magnitude. In ATLAS, changes in τ_{eff} are accounted for automatically, provided relevant recombination models and their associated parameters are correctly implemented. This step is crucial since the IV characteristics of the modulator very much depend on carrier recombination.

Four fundamental recombination mechanisms are known to affect τ_{eff} in Ge: SRH and surface recombination in which we have some degree of control; and radiative and Auger recombination, both intrinsic properties of the crystal.

In SRH recombination, transition metals have been identified as the most efficient lifetime 'killer' in Ge even at small densities (Fig. 4.15). That is why we go to great lengths to protect Ge wafers from metal contamination (e. g., frequently cleaning all beakers and tweezers with HF mixtures, low temperature processing, using high purity

metals for forming metal layers etc). However, not all impurities are harmful. Some are added intentionally to create heavily doped 'n' and 'p' regions on Ge for producing active devices (Sec. 4.2.1). For example, elements such as P, As and Sb are known donors in Ge, while acceptors are usually B, Ga, and Al. These elements are much less detrimental to carrier lifetime because they form trap levels far away from the intrinsic fermi level E_i (see Eq. 4.9), and have cross sections for carrier capture in the order of magnitude less compared to Cu [9]. Comparison between Cu and Ga is shown in Fig. 4.15. Even so, recombination due to donors and acceptors must be accounted for

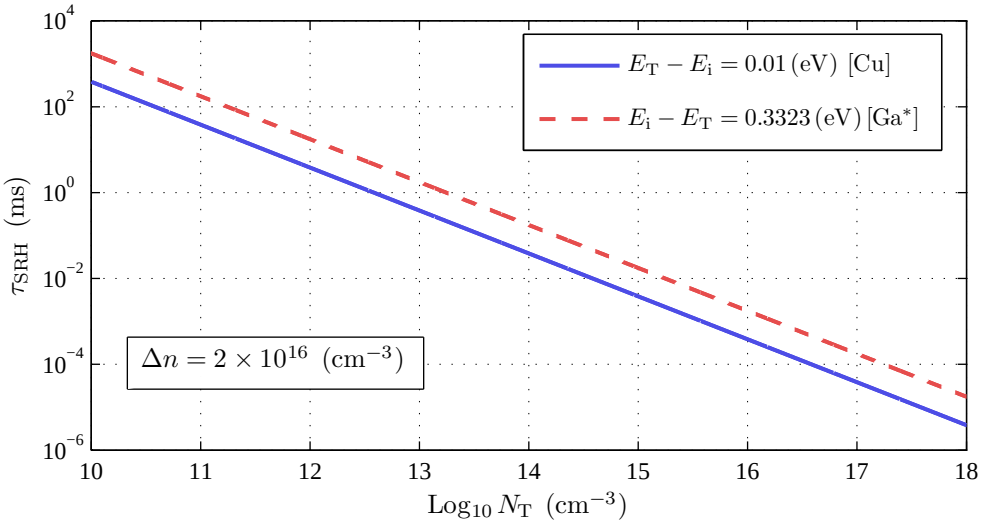


Figure 4.15: τ_{srh} as a function of impurity concentration. SRH parameters for Ga extracted from data published by Gaubas *et. al.* [9]

if they are present at high densities. For us, these take the form of shallow implants beneath the electrodes of the modulator.

For intrinsic Ge, both radiative and Auger recombination dominate over SRH recombination at high carrier densities. Because of the square dependency of Auger recombination on Δn, it quickly surpasses that of radiative, but only for Δn above $5 \times 10^{17}\,\mathrm{cm}^{-3}$. Both however, are detrimental to carrier lifetime at different regions of the modulator. Close to the electrodes, carrier densities can easily exceed $10^{17}\,\mathrm{cm}^{-3}$. Auger recombination thus dominates, and is the main reason for the large forward potential drop across the modulator. Radiative recombination affects mainly carriers that diffuse to the centre of the aperture (away from the annular electrodes). Depending on modulator construction, at 2.6 A of forward current, carriers at densities $\approx 10^{16}$ to $5 \times 10^{16}\,\mathrm{cm}^{-3}$ may flood the aperture—densities where τ_{rad} overshadows other recombination mechanisms (Fig. 4.4). It must be stressed that due to photon recycling,

τ_{rad} may not be as low as we have calculated. This has been argued by Dumke [7] in the early fifties, though many mainstream text on high power device tend to ignore radiative recombination completely without satisfactory justifications given [3, 10].

Because the modulator works by diffusing carriers to the middle of the annular electrodes (Fig. 5.6), diameters exceeding a few L_{a} will result in large reduction in absorption at the middle. Since L_{a} is a function of carrier lifetime, it imposes a fundamental limit to the maximum aperture size. Thus for larger apertures (the default size being 5 mm), new modulator structures must be implemented. One example is to produce electrode mesh across the aperture, allowing carriers to be injected at different locations. This method is undesirable because it will lower the absolute transmission during modulator off state.

Bibliography

[1] R. B. Adler, A. C. Smith, and R. L. Longini. *Introduction to Semiconductor Physics*, volume 1. John Wiley & Sons, New York, 1964.

[2] V. Badescu and P. T. Landsberg. Theory of some effects of photon recycling in semiconductors. *Semiconductor Science and Technology*, 8:1267–1276, 1993.

[3] B. J. Baliga. *Modern Power Devices*. John Wiley and, 1st edition, 1987.

[4] A. Blondeel, P. Clauws, and B. Depuydt. Lifetime measurements on ge wafers for ge/gaas solar cells—chemical surface passivation. *Material Science in Semiconductor Processing*, 4:301–303, 2001.

[5] P. H. Brill and R. F. Schwarz. Radiative recombination in germanium. *Physical Review*, 112(2):330–333, 1958.

[6] R. Conradt. Minority carrier lifetime in highly doped ge. *Solid State Communications*, 10:321–323, 1972.

[7] W. P. Dumke. Spontaneous radiative recombination in semiconductors. *Physical Review*, 105(1):139–143, 1957.

[8] P. D. Fairley. *Novel Solid State Modulator for the Infrared: The Germanium Chopper*. PhD thesis, University of Southampton, 2000.

[9] E. Gaubas and J. Vanhellemont. Comparative study of carrier lifetime dependence on dopant concentration in silicon and germanium. *Journal of The Electrochemical Society*, 154(3):H231–H238, 2007.

[10] S. K. Ghandi. *Semiconductor Power Devices*. John Wiley & Sons, 1977.

[11] A. Haug. Auger recombination of electron-hole drops. *Solid State Communications*, 25:477–479, 1978.

[12] J. R. Haynes and W. Shockley. The mobility and life of injected holes and electrons in germanium. *Physical Review*, 81(5):835–843, 1951.

[13] A. Herlet. The forward characteristic of silicon power rectifiers at high current densities. *Solid State Electronics*, 11:717–742, 1968.

[14] N. R. Howard and G. W. Johnson. Pin silicon diodes at high forward current densities. *Solid State Electronics*, 8:275–284, 1965.

[15] L. Huldt. Band-to-band auger recombination in indirect gap semiconductors. *Physica Status Solidi (A)*, 8:173–187, 1971.

[16] V. K. Khanna. Carrier lifetimes and recombination-generation mechanisms in semiconductor device physics. *European Journal of Physics*, 25:221–237, 2004.

[17] J. Linnros. Carrier lifetime measurements using free carrier absorption transients. ii. lifetime mapping and effects of surface recombination. *Journal of Applied Physics*, 84(1):284–291, 1998.

[18] J. Linnros. Carrier lifetime measurements using free carrier absorption transients. i.principle and injection dependence. *Journal of Applied Physics*, 84(1):275–283, 1998.

[19] S. Marchetti, M. Martinelli, R. Simili, M. Giorgi, and R. Fantoni. Measurement of ge electrical parameters by analysing its optical dynamics. *Physica Scripta*, 64:509–511, 2001.

[20] D. Poelman, P. Clauws, and B. Depuydt. Chemical surface passivation of low resistivity p-type ge wafers for solar cell applications. *Solar Energy Materials and Solar Cells*, 76:167–173, 2003.

[21] J. P. Ponpon. Evolution with time the surface properties of high-purity germanium. *Nuclear Instruments and Methods in Physics Research A*, 457:262–265, 2001.

[22] D. K. Schroder. *Semiconductor Material and Device Characterization*. John Wiley & Sons, New Jersey, 3rd edition, 2006.

[23] E. Simoen, C. Claeys, S. Sioncke, J. V. Steenbergen, M. Meuris, S. Forment, J. V. Vanhellemount, P. Clauws, and A. Theuwis. Lifetime and leakage current considerations in metal-doped germanium. *Journal of Material Science: Materials in Electronics*, 18:799–804, 2007.

[24] R. A. Smith. *Semiconductors*. Cambridge University Press, Cambridge, 2^{nd} edition, 1978.

[25] S. M. Sze. *Physics of Semiconductor Devices*. John Wiley & Sons, 2nd edition, 1981.

[26] S. M. Sze and K. K. Ng. *Physics of Semiconductor Devices*. John Wiley & Sons, New Jersey, 3rd edition, 2007.

[27] Y. P. Varshni. Band-to-band radiative recombination in groups iv, vi and iii-v semiconductors(i). *Physica Status Solidi*, 19:459, 1967.

[28] E. Yablonovitch, D. L. Allara, C. C. Chang, T. Gmitter, and T. B. Bright. Unusually low surface recombination velocity on silicon and germanium surfaces. *Physical Review Letters*, 57(2):249–252, 1986.

[29] J. F. Young and H. M. van Driel. Ambipolar diffusion of high-density electrons and holes in ge, si, and gaas: Many body effects. *Physical Review B*, 26(4):2147–2158, 1982.

Chapter 5

Device Simulation

Having reviewed the parameters that influence the performance of the modulator through the use of empirical equations (chapters 3 and 4), results from utilizing a far superior method for predicting the electrical properties of the modulator are presented in this chapter. These are produced using ATLAS, a physically-based device simulator that is designed to specifically model the behaviour of semiconductor devices. Its advantage lies in being able to simulate three-dimensional devices using complex numerical solvers; thus giving better predictions of the behaviour of real devices. Included also are results from modifications to the default modulator structure and its biasing conditions, which can be employed to augment device performance. For completeness, the chapter starts off with an overview of ATLAS and its general capabilities. Steps involved in setting up a simulation model are then described, which include additional information on the implementation of carrier mobility and recombination models in ATLAS not covered in chapter 3. (All materials related to the use of ATLAS were summarized from the ATLAS manual [20]). Finally, results from simulation work conducted are presented.

5.1 The Nature of Physically-Based Device Simulation

Unlike empirical modelling which approximates existing data using analytical equations to acceptable accuracy, a physically-based device simulator employs a variety of elaborate and powerful numerical solvers to predict the properties of a structure under different stimulus. In our case, a PiN diode subjected to different biasing conditions is simulated. Provided that appropriate physical models are invoked (e. g., semiconductor bandgap, carrier mobility and carrier recombination mechanisms), and boundary conditions specified (e. g., electrode contact and semiconductor surface properties),

the electrical behaviour of devices with different shape, size and construction can be predicted to a good accuracy without experimenting with actual devices. Device simulation is therefore a major part of the PhD, taking up to a third of the entire duration since it is a quicker and cheaper method for optimizing the design of the modulator compared to performing experiments.

ATLAS works by first providing it with a description of a device meshed with a two or three dimensional grid using ASCII command line instructions. These grids contain a number of grid points called nodes that are limited to a maximum of $20,000$, a constraint set by ATLAS. With appropriate numerical solvers invoked (by the user), ATLAS then applies a set of differential equations onto the grid (section 5.2). With a bias profile specified, the properties of carriers in the device can then be solved through an iterative procedure. Devices can be modelled in direct and alternating current (DC and AC) settings, or under transient mode of operation. The user has to provide the structural information of a device to be simulated (with appropriate mesh defined), invoke the appropriate physical models and their associated numerical solvers, and to set the desired bias profile for ATLAS to predict the electrical behaviour of a particular device.

5.2 Basic Equations

The static and dynamic behaviour of carriers in semiconductor devices are described using three fundamental equations: the electrostatic equation, the current-density equation and the continuity equation [22]. With a simulation model in place and the bias specified, ATLAS solves all three equations for the user automatically. However, the three-dimensional form of these equations are briefly reviewed in this section because they give insight into the behaviour of carriers, which helps in efforts to optimize device performance.

The electrostatic equation relates the local charge densities to the electric field. Written as

$$\nabla \cdot \boldsymbol{E} = \frac{1}{\epsilon_{\mathrm{s}}} \rho(x, y, z), \tag{5.1}$$

where $\boldsymbol{E}$ is the electric field (V/cm), ϵ_{s} the semiconductor permittivity (F/cm), and ρ the charge density (C/cm^3). By integrating equation 5.1, the electric field distribution inside a semiconductor (caused by the build-up of charges) can be evaluated [7, 20]. (All vectors are written using boldface italic.) Depending on the magnitude of $\boldsymbol{E}$, different

mobility models must be used to accurately model the behaviour of carriers as described in Sec. 3.3.5. The low field mobility approximation was assumed in ATLAS since electric fields caused by *excess carriers* do not exceed 15 V/cm in our device [10, 15]. However, within and in the vicinity of the doped region, the local electric fields can be very high (>1000 V/cm) due to the presence of ionized dopants. These regions extend no more than a few micrometers from the electrode thus it should not affect the performance of the modulator in the scale that is of interest to us.

The current density equations give the total electron and hole current densities as the sum of drift and diffusion components. The drift component is driven by electric field, while the diffusion component is caused by the gradient in carrier concentration [22]. The current density equations for electrons and holes are given as:

$$\boldsymbol{J}_{\mathrm{n}} = q\mu_{\mathrm{n}} n\boldsymbol{E} + qD_{\mathrm{n}}\nabla n \tag{5.2}$$

$$\boldsymbol{J}_{\mathrm{p}} = q\mu_{\mathrm{p}} p\boldsymbol{E} + qD_{\mathrm{p}}\nabla p. \tag{5.3}$$

$\boldsymbol{J}$ is the current density ($\mathrm{A/cm^2}$), n and p are the concentration of free electron and holes respectively ($\mathrm{cm^{-3}}$), μ the drift mobility of carriers ($\mathrm{cm^2/V \cdot s}$), q the electronic charge (C), and D is the diffusion coefficient of carriers ($\mathrm{cm^2/s}$). The application of the current density equations in one-dimension have been demonstrated in Sec. 4.2.1. Of interest to us is the diffusion component because it is the mechanism responsible for literally flooding the modulator with IR absorbing carriers. As can be seen from equations 5.2 and 5.3, its magnitude is proportional to the gradient of non-equilibrium free carrier concentration. In our device, carrier diffusion is caused by the build-up of excess carriers drifting between the contacts during forward bias (figure 5.6).

While the current density equations are for steady state conditions (i. e., the new equilibrium state of a system acted upon by constant external forces), the continuity equation deals with the evolution of carriers with time, encompassing time dependent mechanisms such as carrier injection, and carrier generation and recombination. The continuity equations for electrons and holes are defined as [22]:

$$\frac{\partial n}{\partial t} = G_{\mathrm{n}} - R_{\mathrm{n}} + \frac{1}{q}(\nabla \cdot \boldsymbol{J}_{\mathrm{n}}) \tag{5.4}$$

$$\frac{\partial p}{\partial t} = G_{\mathrm{p}} - R_{\mathrm{p}} + \frac{1}{q}(\nabla \cdot \boldsymbol{J}_{\mathrm{p}}). \tag{5.5}$$

G and R are the generation and recombination rates of carriers ($\mathrm{cm^{-3} \cdot s^{-1}}$). When excess carriers are injected into the modulator, the generation term becomes negligible as the concentration of carriers far exceeds thermal equilibrium values ($np \gg {n_{\mathrm{i}}}^2$) [11].

The density of electrons and holes that flood the modulator during forward bias are thus determined by the difference between R and the diffusion component of $\boldsymbol{J}$.

5.3 Simulation Model Set-up

The first step in setting up a simulation model is to define the structure of the modulator. This is fed into ATLAS via an input file containing command line instructions. (Device structures can be created using software that works in conjunction with ATLAS such as DEVEDIT and ATHENA, or it can be entered via command line instructions. We have used the second method as it is the quickest way to define simple structures and to perform modifications to it.)

To aid illustration, the general structure of the modulator specified in ATLAS is shown in figure 5.1. As can be seen, only a segment of the cross section of the modulator is needed for ATLAS to perform three-dimensional (3D) computations on the modulator. This was possible due to the cylindrical structure of the device, which enabled the use of the function `cylindrical`[1] to prompt ATLAS to produce results equivalent to a full 3D model of the modulator. Once activated, the centre of the aperture is set as the axis of symmetry around which a cylindrical structure can be assumed for simulation. Similarly, results can only extracted from the user specified 2D structure, since ATLAS does not generate a 3D model of the device for analysis. For this reason, maps illustrating different properties of the modulator in subsequent sections are all in 2D.

For the remaining part of the set-up (except for result analysis), command line instructions must be used to instruct ATLAS to solve a simulation problem. The order in which these instructions occur is important, and can be divided into five groups of statements outlined in figure 5.2. Failure to adhere to this order may result in error messages being produced or even a premature termination of a simulation run. Some statements may also be ignored by ATLAS, producing erroneous results.

The order of statements *within* each group such as the structural definition, model specification, and solution groups is also important. Failure to place these statements in the proper order may result in the aforementioned complications. For the most part, it is a fairly straight forward procedure to define a simulation structure and to invoke the relevant physical models for simulation. Users are referred to the ATLAS manual for

[1] ATLAS syntax are written in `Typewriter family font`

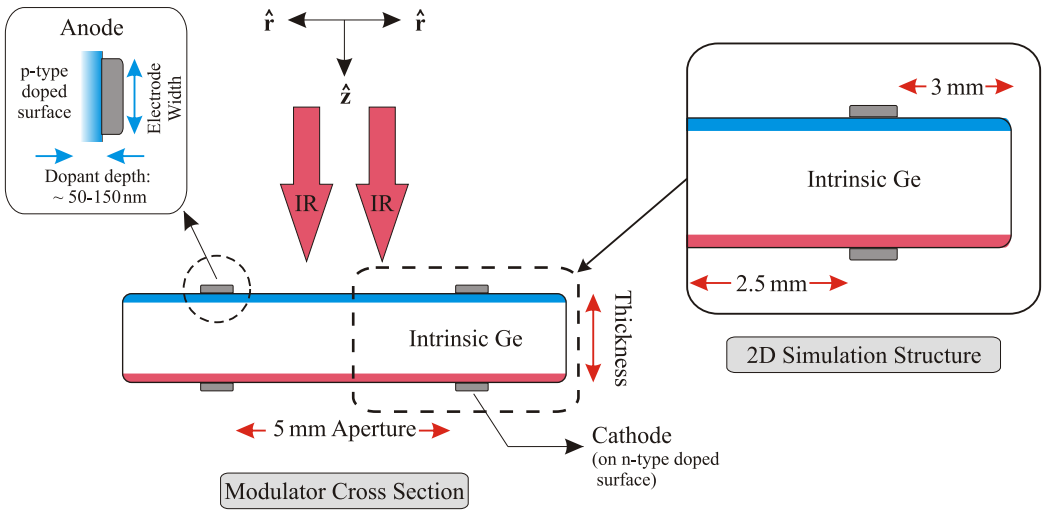


Figure 5.1: The cross section of the modulator and the two-dimensional structure needed by ATLAS to perform computations in three dimensions. (The diagram uses the cylindrical coordinate system and is not drawn to scale.)

further information [20]. However, important procedures such as the meshing of device structure and the implementation of SRH recombination model 4.1.3 are elaborated in this section because they have significant influence on simulation results.

5.3.1 Structure Meshing

When generated using command line instructions, a mesh is defined by a series of horizontal and vertical lines. A good mesh must be specified for proper simulations to be conducted; too dense a mesh will result in long computation times, while inaccurate/wrong predictions will be produced if it is too coarse. Hence there is a trade-off between the requirements of accuracy and numerical efficiency, offset by the way the mesh is defined. In regions where changes occur rapidly such as doped regions with graded profiles, or under the electrodes where the gradient of the electric field is steep, a denser mesh must be allocated. The time needed for a simulation run to complete is roughly proportional to N^{a}, where N is the number of nodes (grid points) and 'a' is a constant that varies from 2 to 3 depending on the complexity of the problem. Unfortunately, the size of our device pushes the number of nodes used close to $20,000$—the maximum number of nodes permitted by ATLAS. The meshing procedure thus concerns the way nodes are distributed throughout the device.

A significant number of device simulation problems can be avoided by adhering to the following principles. [20]:

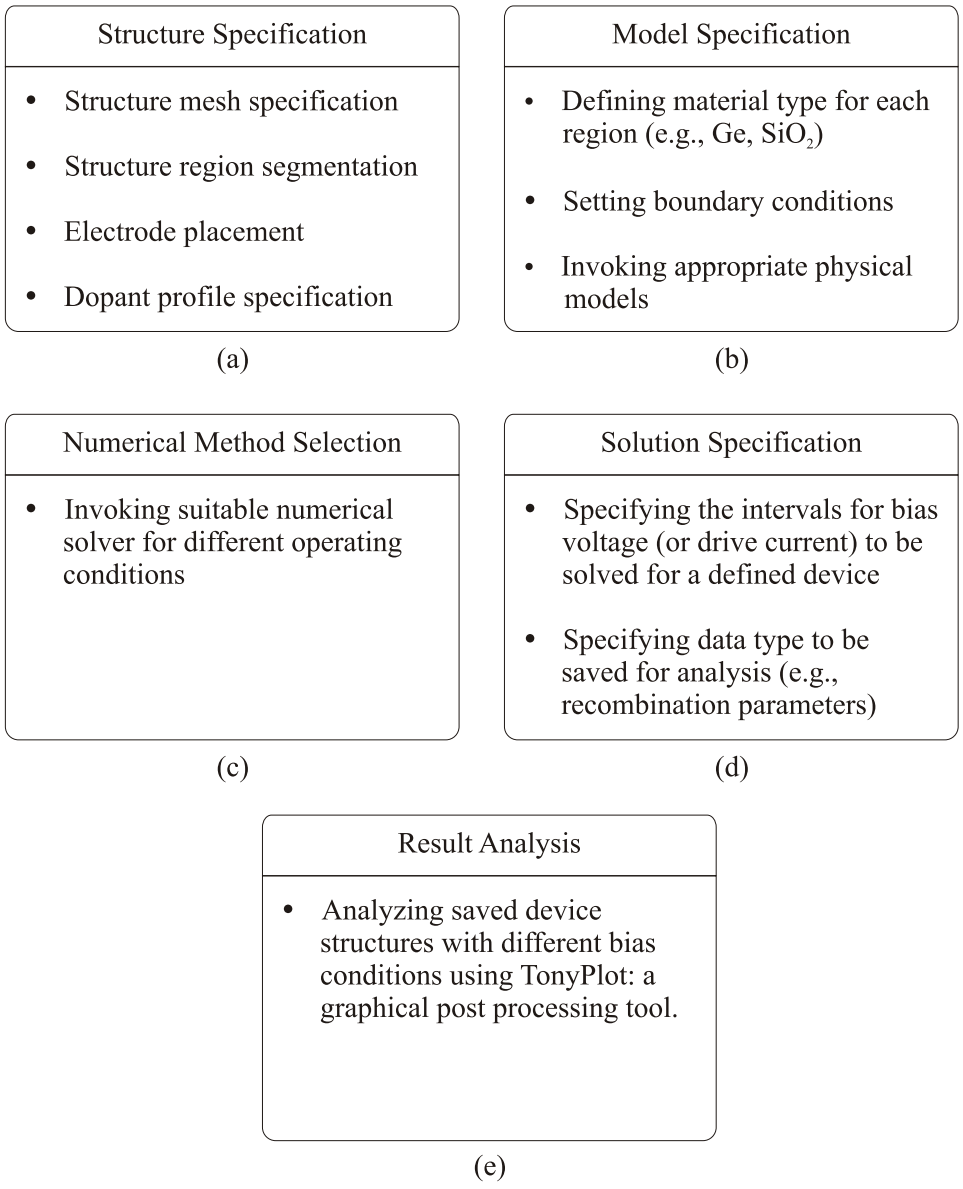


Figure 5.2: The order for each command groups to be specified in ATLAS: from (a) to (e).

1. avoid obtuse triangles in semiconductor regions, particularly in current path and high field areas.

2. avoid abrupt discontinuities in mesh density.

3. avoid "thin" triangles, where the ratio of longest to shortest edge in a mesh triangle should be on the order of 10, but not 100.

4. adjust the mesh density for layered materials, particularly for very thin material sandwiched between thick layers composed of different materials.

The structure of the modulator with mesh assigned is shown in figure 5.3a.

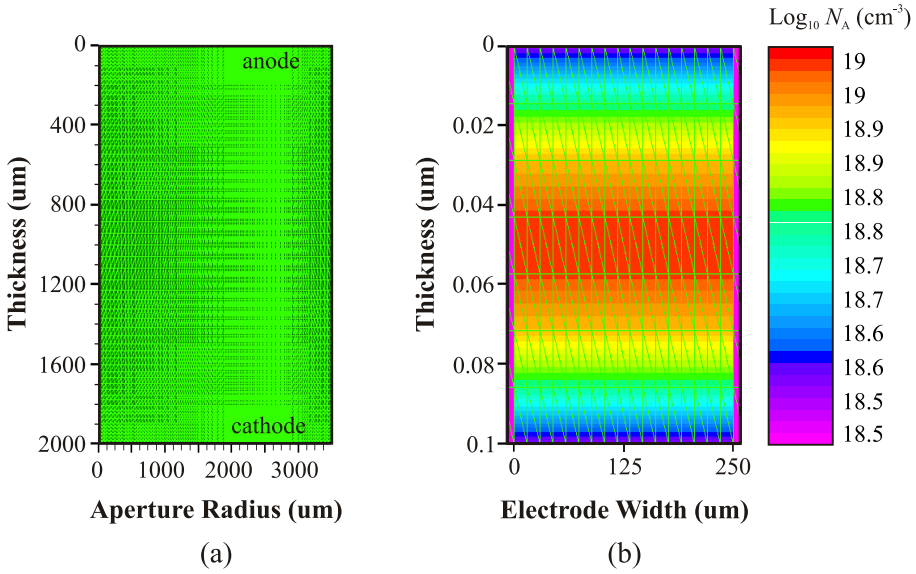


Figure 5.3: (a) The structure of the modulator with varying mesh densities; (b) An example of mesh spacings allocated for the layer beneath the anode doped with acceptors.

5.3.2 Simulation Parameters

The default parameters used for simulating the electrical properties of the modulator are shown in table 5.1. The dependence of carrier mobility on lattice temperature, carrier concentration and dopant concentration have been reviewed in chapter 3, while recombination mechanisms relevant to the functionality of the modulator were detailed in chapter 4. The structural parameters of the modulator were modified from initial simulation work carried out by a former colleague in our group[2]; recombination param-

[2]Suresh Uppal, a former post doctoral research fellow in our group produced the first unpublished *basic* steady state model of the modulator. Although optimistic values were assigned to certain param-

eters assigned with values that err to the side of pessimism; and mobility parameters that best reflect the behaviour of carriers in germanium were chosen.

5.3.3 Ion Implantation and Modulator Doping Profile

As discussed in section 4.2.1, regions beneath the electrodes of the modulator must be doped to obtain good current modulation. The conventional method for achieving this without exposing the modulator to prolonged high temperature processing is by ion implantation [27]. Nevertheless, without performing preimplantation steps to increase the density of electrically active dopants(e. g., surface preamorphization to increase the solid solubility limit of dopants embedded upon recrystallization [2]), or causing severe infrared (IR) attenuation between $8 - 14\,\mu$m by using high implant doses, only shallow layers ($< 150\,$nm) can be doped with electrically active ions to concentrations above $10^{19}\,\mathrm{cm}^{-3}$ [4, 17]. (Because the profile of implanted ions can be approximated by a Gaussian function with fixed characteristics, larger implant doses must be used to achieve similar peak concentration at greater depths. This increases the number of carriers in the modulator, which in turn decreases the transmission through the modulator when it is in the electrically "off" state.)

The doping profiles for the 'n' and 'p' layers under the electrodes were similarly defined in ATLAS. They are shown in figure 5.4. For comparison, the actual concentration profile of donor (phosphorus) and acceptor (boron) ions implanted on the surface of the modulator are included. The spreading resistance method (SRP) was used to determine the density of electrically active dopants (section 6.7.5). The concentration of dopants profiled using SRP may be slightly overestimated because no calibration standards were available for germanium at time of testing [3, 21]. This is discussed in further detail in section 6.7. Nevertheless, the difference between the profile specified in ATLAS and the actual profile measured using SRP does not present a cause for concern because the modulator actually performs better at higher dopant densities (section 5.4.2). The profile in ATLAS was deliberately set to reflect the performance of a pessimistic model.

eters such as carrier lifetime, it provided insight into the behaviour of the modulator [25] which greatly helped in the creation of the present simulation model. We greatly acknowledge the contribution of his pioneering work.

Default Structure Specifications

	Value	Units	Distribution
Wafer Doping	1×10^{10} (n-type)	cm^{-3}	Constant
Peak Dopant Density	1×10^{19} (p-type)	cm^{-3}	Gaussian
	1×10^{19} (n-type)	cm^{-3}	Gaussian
Temperature	300	K	—
Aperture Diameter	5	mm	—
Wafer Thickness	2	mm	—
Electrode Width	250	μm	—
Contact Resistance	0	Ω	—

Default Recombination Parameters

	Parameter	Units	Ref.
SRH Recombination	$\tau_p = 1.6$	s	Sec. 4.1.3
	$\tau_n = 16$	s	
	$\lvert E_T - E_i \rvert = 0.0023$	eV	
	`NSRHN` $= 10^9$	cm^{-3}	Sec. 4.1.3
	`NSRHP` $= 10^9$	cm^{-3}	
Radiative Recombination	$B = 5.25 \times 10^{-14}$	cm^3/s^{-1}	Sec. 4.1.4
Auger Recombination	$C_p = 1 \times 10^{-31}$	cm^6/s	Sec. 4.1.5
	$C_n = 1 \times 10^{-31}$	cm^6/s	
Surface Recombination	SRV = 300	cm/s	Sec. 4.1.6

Default Carrier Mobility Parameters

Parameters	Electrons	Holes	Units	Ref.
μ_{max}	4100	1800	$cm^2/V \cdot s$	Sec. 3.3
μ_{min}	270	50	$cm^2/V \cdot s$	
N_{ref}	7×10^{16}	1.5×10^{17}	cm^{-3}	
α	0.636	0.515	—	
`ALPHA.CAUG`	0	0	—	
`BETA.CAUG`	−1.66	−2.33	—	
`GAMMA.CAUG`	−0.7	−1.7	—	

Table 5.1: The default simulation parameters used in ATLAS. Unless stated otherwise, all simulation results were based on parameters listed in this table.

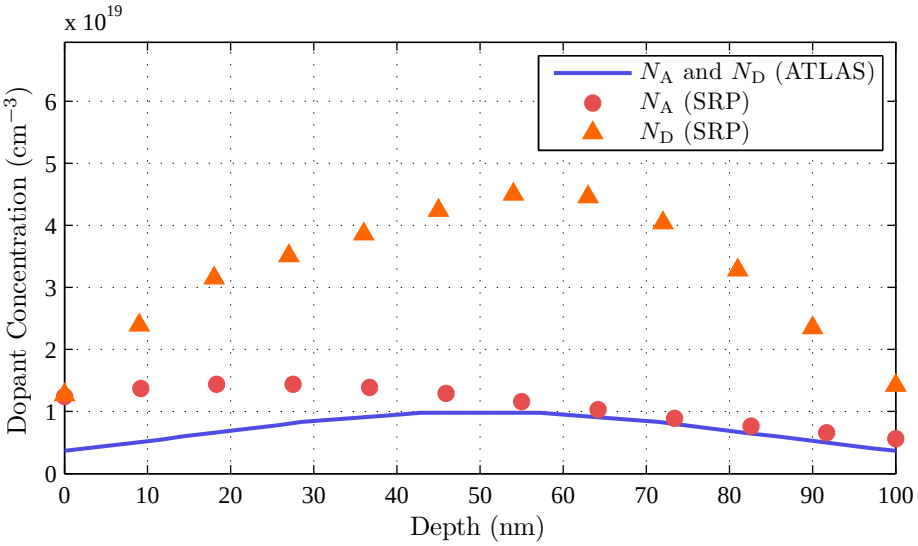


Figure 5.4: A comparison between the doping profile defined it ATLAS for the 'n' and 'p' layer (figure 5.1) and the actual concentration of donor and acceptor ions implanted on the surface of the modulator.

5.3.4 Implementation of Shockley-Reed-Hall Recombination

Because impurities act as recombination centres for free carriers, SRH recombination (section 4.1.3) must be implemented in ATLAS for the current density equations to be solved accurately (equations 5.2 and 5.3). If the doping level is constant, equation 4.9 can be used to evaluate the lifetime of carriers due to SRH recombination. But for dopants with graded profiles, the electron and hole lifetime parameters found in equation 4.1.3 must be replaced to account for changes to the doping level. Expressed as

$$\tau_{\mathrm{n}} = \frac{1}{\sigma_{\mathrm{n}}\,\nu_{\mathrm{th}}\,N_{\mathrm{T}}} \tag{5.6}$$

$$\tau_{\mathrm{p}} = \frac{1}{\sigma_{\mathrm{p}}\,\nu_{\mathrm{th}}\,N_{\mathrm{T}}}, \tag{5.7}$$

for constant doping level, they must be replaced with the following empirical equations for devices with non-uniform doping profiles [16, 20]:

$$\tau_{\mathrm{n}}^{*} = \frac{\tau_{\mathrm{n}}}{1 + N_{\mathrm{D}}/\mathtt{NSRHN}} \tag{5.8}$$

$$\tau_{\mathrm{p}}^{*} = \frac{\tau_{\mathrm{p}}}{1 + N_{\mathrm{A}}/\mathtt{NSRHP}} \tag{5.9}$$

where σ_{p} and σ_{n} are respectively the cross sections for hole and electron capture by donors and acceptors (cm^{-2}); N_{D} and N_{A} the corresponding doping concentration

(cm^{-3}); and ν_{th} the thermal velocity expressed as $\frac{1}{2}m^*\nu_{\mathrm{th}}^2 = 3/2kT$. Parameters τ_{n}, τ_{p}, `NSRHN` and `NSRHP` are user defined parameters. It turns out that when τ_n and τ_p are set, the values for `NSRHN` and `NSRHP` are unique. The opposite is true if `NSRHN` and `NSRHP` are first defined. Therefore, so long as equations 5.8 and 5.9 agrees with equations 5.6 and 5.7 for different levels of fixed dopant density, the recombination model should behave as intended. This is illustrated in figure 5.5, where the solid line represents the actual lifetime of electrons.

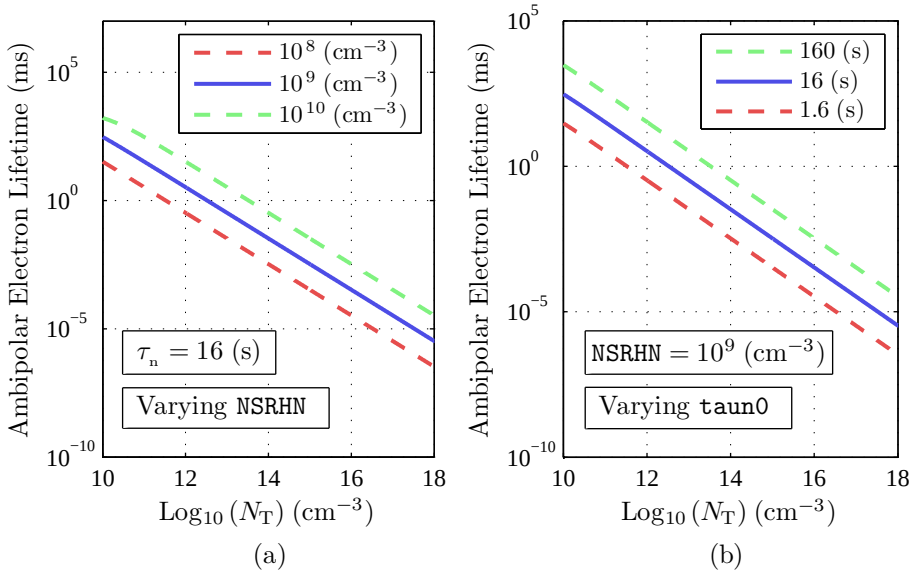


Figure 5.5: The electron lifetime for fixed (a) τ_n, and (b) `NSRHN`. The correct values are plotted using solid lines, highlighting the dependence between τ_{n} and `NSRHN`.

5.3.5 Modulator Performance

As carriers are injected into the modulator during forward bias, they form drift currents that flow through a small volume between the contacts. At the same time, a small fraction of these carriers will diffuse laterally to flood the entire 'i' region of the modulator because of the gradient in carrier concentration formed by drifting carriers (figure 5.6). Using TONYPLOT, a graphical post processing tool for ATLAS, this process is depicted in figure 5.6.

The infrared transmission through the modulator has to be evaluated manually. This is done using equation 2.25, which gives the absolute transmission as a function of electron and hole area concentration (A_{e} and A_{h} (cm^{-2})). The values for A_{e} and A_{h} are obtained by integrating the density of carriers through the entire depth of the

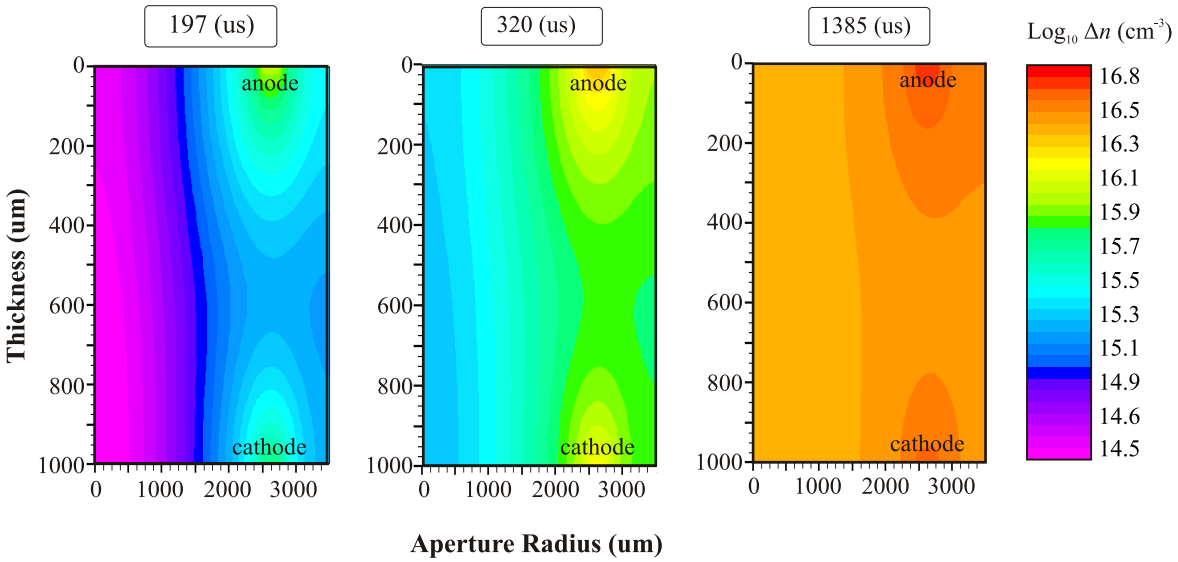


Figure 5.6: The evolution of carriers injected into modulator at 0.5 V applied voltage (V = 0 at T = 0; V = 0.5 at T > 0).

modulator. For convenience, the cross section for carriers at 10.59 μm is used. (Since the number of electrons and holes injected into the modulator are similar, only the area concentration for either type of carriers is needed to calculate the total absorption induced by *both types* of carriers. Therefore, only A_{h} is used in discussions, but the actual transmission plotted in graphs already includes the contribution from electrons and holes) By default, the transmission through the modulator is evaluated at the middle of the aperture, since IR attenuation is the weakest at that point (figure 5.6).

The transmission through the modulator at different values of steady state current flow, and their corresponding power dissipation are plotted in figure 5.8. The default structure requires less than 2 watts of power to achieve $\approx$ 95% IR attenuation. The exponential dependence of equation 2.25 leads to the rapid decay in IR transmission at higher carrier concentration.

The time dependent behaviour of the modulator can be simulated using several time based bias functions including the piecewise-linear, exponential, and sinusoidal functions. Using linear ramping, the transient behaviour of the modulator was simulated using the following profile: (1) linearly ramping the forward voltage to 0.5 V in 10 μs, (2) maintaining the voltage at 0.5 V for 4.5 ms, (3) ramping down the voltage to zero in 10 μs, and (4) leaving the device to reach equilibrium for 4.5 ms. The resulting current through the modulator and the attenuation induced are shown in figure 5.9.

By defining the raise and fall time as the time required to reduce the transmittance

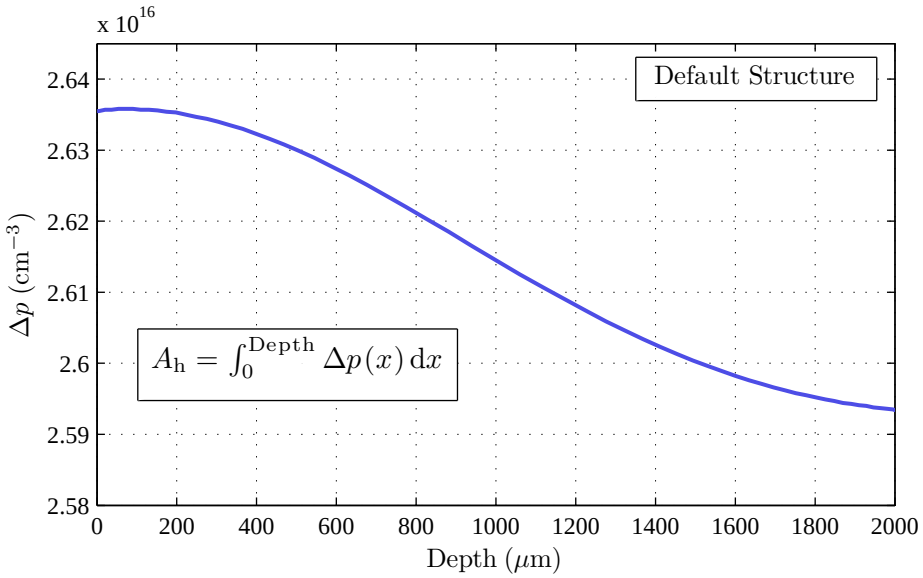


Figure 5.7: The concentration of holes through the middle of the modulator as a function depth. A_h is evaluated by integrating Δp through the entire depth of the modulator.

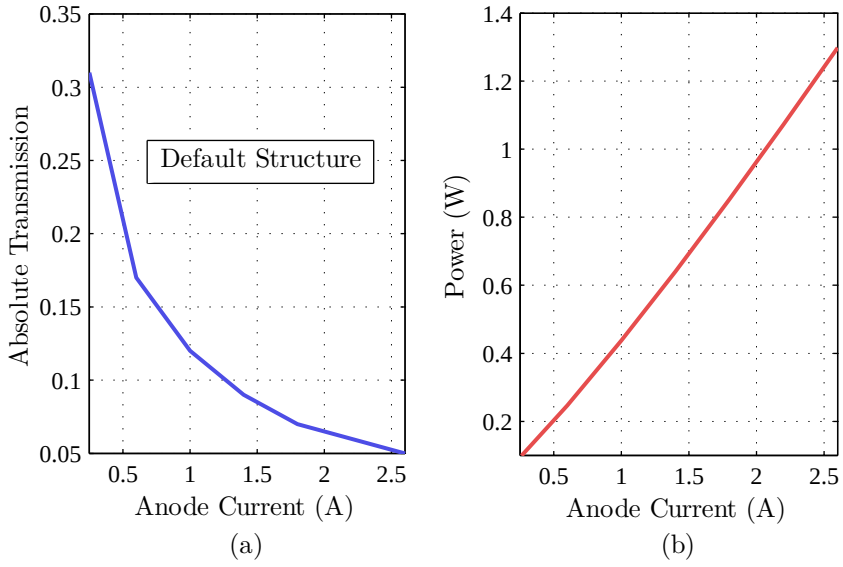


Figure 5.8: (a) Transmission through the modulator for different rates of steady state current flow, and (b) the corresponding power consumption.

from 0.9 to 0.1 and from 0.1 to 0.9 respectively, the default modulator is predicted to operate at frequencies up to at least 200 Hz. The modulator is therefore suited for use in thermal detectors/imagers utilizing pyroelectric detectors, where signal chopping frequencies are less than 50 Hz. The negative spike in current occurs when voltage is reduced to zero. This behaviour is due to device inductance and capacitance [23] as it recovers (i. e., the time needed for excess carriers decay to room temperature intrinsic values).

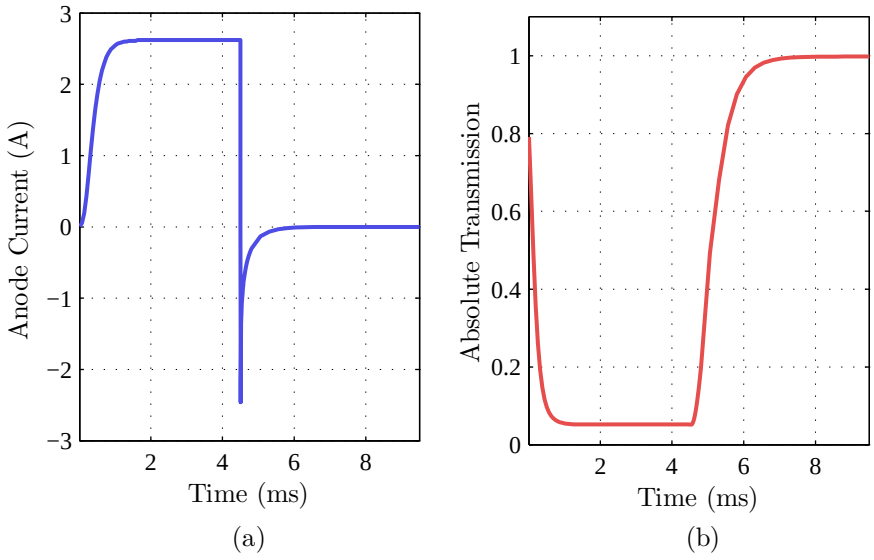


Figure 5.9: The time dependent performance of the modulator.

5.4 Device Optimization: Structural Modifications

Two aspects of the modulator can be optimized: the depth of modulation (modulation strength) and the operating frequency. Guided by the current density equation, modifications to the structural parameters of the modulator have been performed to achieve the former, while the incorporation of additional steps to the bias profile, such as forward voltage overshoot and transient reverse voltage on switch-off to assist carrier removal have been explored to improve the latter.

Two mechanisms are responsible for the distribution and annihilation carriers in the bulk of the modulator (i. e., volume under the aperture): carrier diffusion and recombination. (Drift carriers have little effect on the number of carriers that flood the modulator because they only flow through a small volume *between* the electrodes and

not the aperture). By increasing the rate of the former and/or by decreasing the rate of the latter, the depth of modulation can be improved. Using the default structure as the starting point, structural modifications that can be employed to enhance the rate of carrier diffusion are first presented.

5.4.1 Electrode Width

By increasing the current density through the modulator to produce steeper carrier concentration gradients, the rate of lateral carrier diffusion can be enhanced. This will result in greater depth of modulation as more carriers will flood the modulator before recombining. Without raising the total current through the modulator, the current density can be increased by using narrower electrode widths (figure 5.1). The increase in the hole area density (A_h) using this method is shown in figure 5.10a. A steady state current of 2.6 A was used. However, the rise in A_h is accompanied by an increase in power dissipation. This is caused by current crowding effects at the vicinity of the contacts, where larger voltages must be applied to maintain the same rate of carrier flow through electrodes with smaller contact areas [19].

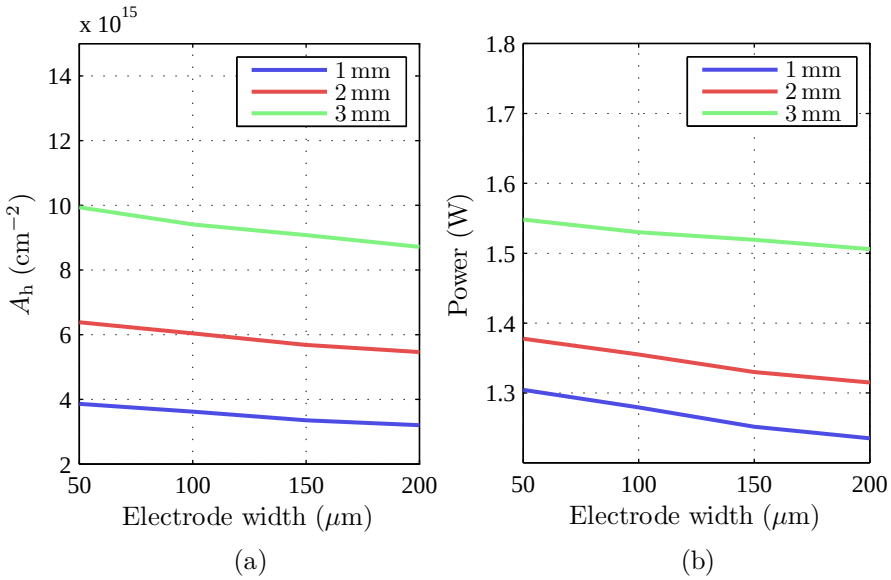


Figure 5.10: (a) The dependence of hole area density (A_h) as functions of electrode width and modulator thickness at 2.6 A, and (b) the power dissipated by the modulator.

Because a large fraction of injected carriers will drift through the device without contributing much to modulator performance, it is important that narrow electrodes be used to reduce power dissipation. Although this requirement contradicts the choice of width used in the default structure (standard lithographic processing could easily achieve 10 μm resolution), it was deemed best for initial fabrication work in view of our

inexperience in device processing. The third revision of the modulator was successfully produced using 50 μm wide electrodes.

As the density of carriers increases, so too will the rate of Auger recombination, which is proportional to the square of the carrier density. Thus improvements to the depth of modulation using higher drift current densities will plateau at some point, as the effect of increased carrier diffusion is negated by Auger recombination.

5.4.2 Peak Implant Concentration

Another parameter that greatly affects the performance of the modulator is the density of dopants in the 'n' and 'p' layers (figure 5.1). A high doping density is needed to increase the efficiency of carrier injection, which works by: (a) reducing the *width* of potential barrier faced by carriers entering and leaving the modulator (section 3.4.3), and (b) lowering the number of electrons and holes from being injected into the 'p' and 'n' layer respectively—both of which do not contribute to current modulation during forward bias (section 4.2.1). ATLAS has shown that such layers can be as thin as 20 nm, and it is the peak dopant concentration that significantly affects the electrical behaviour of the modulator. This is illustrated in figure 5.11, where A_{h} is simulated as a function of peak dopant density. As the doping level drops to intrinsic value, the PiN structure of the modulator essentially becomes a resistor, evident from the increase in power dissipation (figure 5.11b).

It is well known that the solid solubility limit for each species of dopants in germanium are different [24]. Achieving this limit, however, depends on the method by which dopants are incorporated. More importantly, a large fraction of the dopants must be electrically active for efficient modulators to be produced. As alternative methods to cheaply and effectively dope germanium was tried (section 6.8.3), constraints on the maximum allowable processing temperature meant that the density of electrically active dopants could vary by few orders of magnitude, brought about by the nature in which dopants are incorporated. It is therefore vital that the behaviour of the modulator with asymmetric 'n' and 'p' doped layers be simulated (i. e., having dissimilar doping density). With the profile of the 'p' layer unchanged, the predicted performance of the modulator with varying peak donor concentration (N_{D}) is shown in figure 5.12. At low currents, the power dissipated is essentially the same for all doping levels because enough *excess* carriers are present to modulate current flow (i. e., form the majority of carriers that drift through the modulator). At higher currents, the drawbacks of low

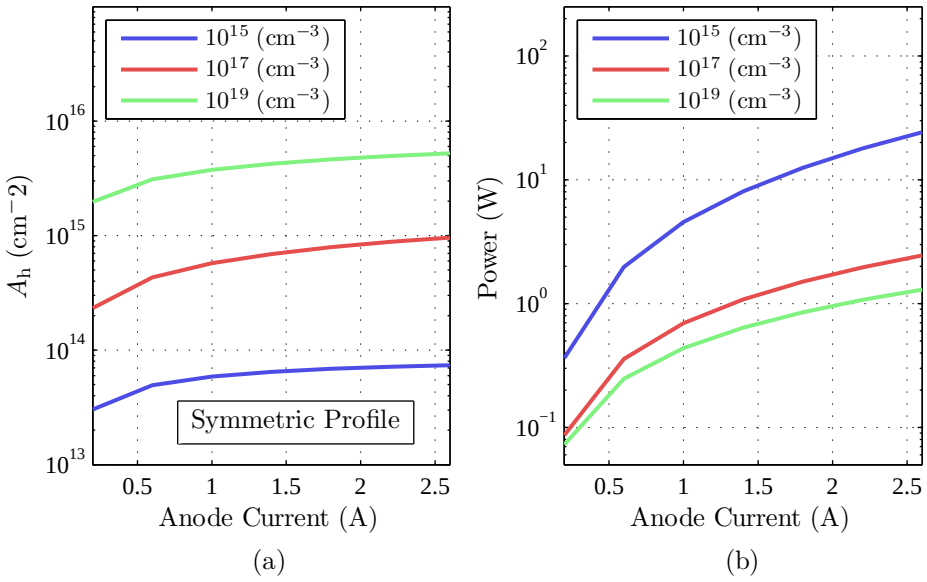


Figure 5.11: The resulting hole area concentrations as a function of peak dopant concentration.

peak doping density become evident as more power is needed to produce a similar flow of current. This shows that less carriers are injected into the modulator. As a result, A_h drops by roughly two orders of magnitude when the peak dopant density is reduced from 10^{19} to $10^{15}\,\mathrm{cm}^{-3}$ at 2.6 A of forward current. (The number of carriers diffusing across the modulator is limited by the charge neutrality requirement. If less electrons are injected, as predicted when the doping density in the 'n' layer is reduced, so too are the number of holes.)

5.4.3 The Influence of Carrier Recombination on Device Performance

As highlighted by the current density equation (equations 5.2 and 5.3), if the recombination rate of carriers is similar to the rate at which excess carriers are introduced into the bulk of the modulator via diffusion (not by thermal generation, G), no modulation can be achieved! Therefore, methods to reduce surface and bulk recombination were vigorously pursued, since the feasibility of an electrically operated modulator hinges on how well recombination mechanisms can be suppressed. To aid discussions, the influence of carrier recombination on the performance of the modulator are first reviewed.

At high carrier densities, most carriers will recombine via radiative and Auger recombination (chapter 4). Although Auger recombination is often assumed to be the

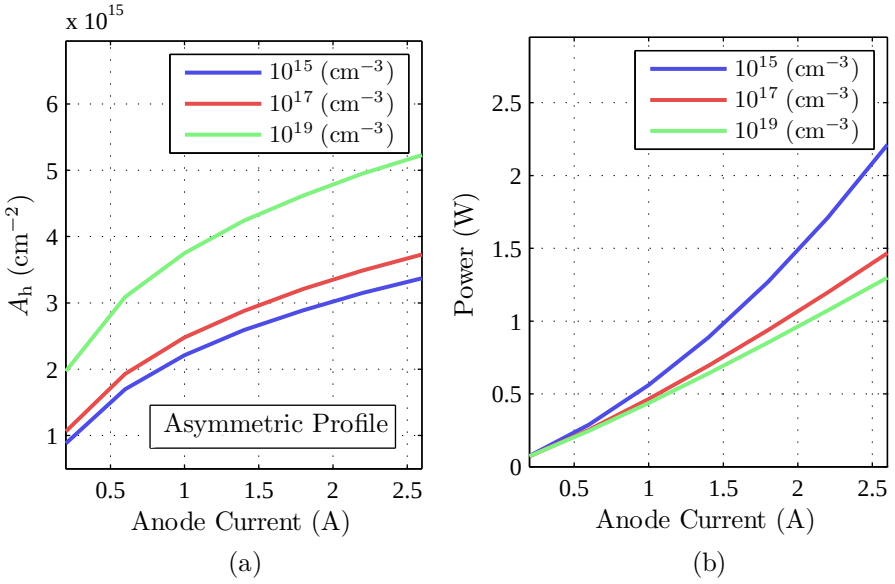


Figure 5.12: (a) The hole area density as a function of different peak donor density in the 'n' doped layer. The profile 'p' layer is unchanged. (b) The increase in power dissipation as a function of decreasing donor concentration.

dominant recombination mechanism in devices that operate at high carrier densities [6, 19], the rate of radiative recombination is higher within the operating domain of the modulator, where the density of carriers range from 3×10^{14} to $3 \times 10^{17}\,\mathrm{cm}^{-3}$ (section 4.1.4). This is illustrated in figure 5.13. (The dominance of radiative recombination at low to medium levels of carrier density ($< 10^{17}\,\mathrm{cm}^{-3}$) is however dependent on the rate of SRH recombination. In certain cases, SRH recombination may even overshadow the effects of radiative recombination—especially when impurities such as copper and nickel are present in large quantities. See section 4.1.4 for discussions on why radiative recombination is prevalent in germanium.)

So far, in the interest of highlighting the mitigating effects of Auger and radiative recombination, it has been assumed that the bulk of the modulator is made up of intrinsic germanium. This was specified to keep SRH recombination to a minimum—one of the two recombination mechanisms that can be controlled through careful device processing. If the intrinsic region of the modulator is contaminated with transition metals (TM) such as copper, only 1 part per billion TM atom is sufficient to render the modulator useless. The modulator does however have a higher tolerance for shallow dopants such as phosphorus, since its carrier capture cross section is roughly ten times lower compared to copper. The deleterious effects of incorporating copper and phosphorus ions in the modulator are compared in figure 5.14. As can be seen, the highly

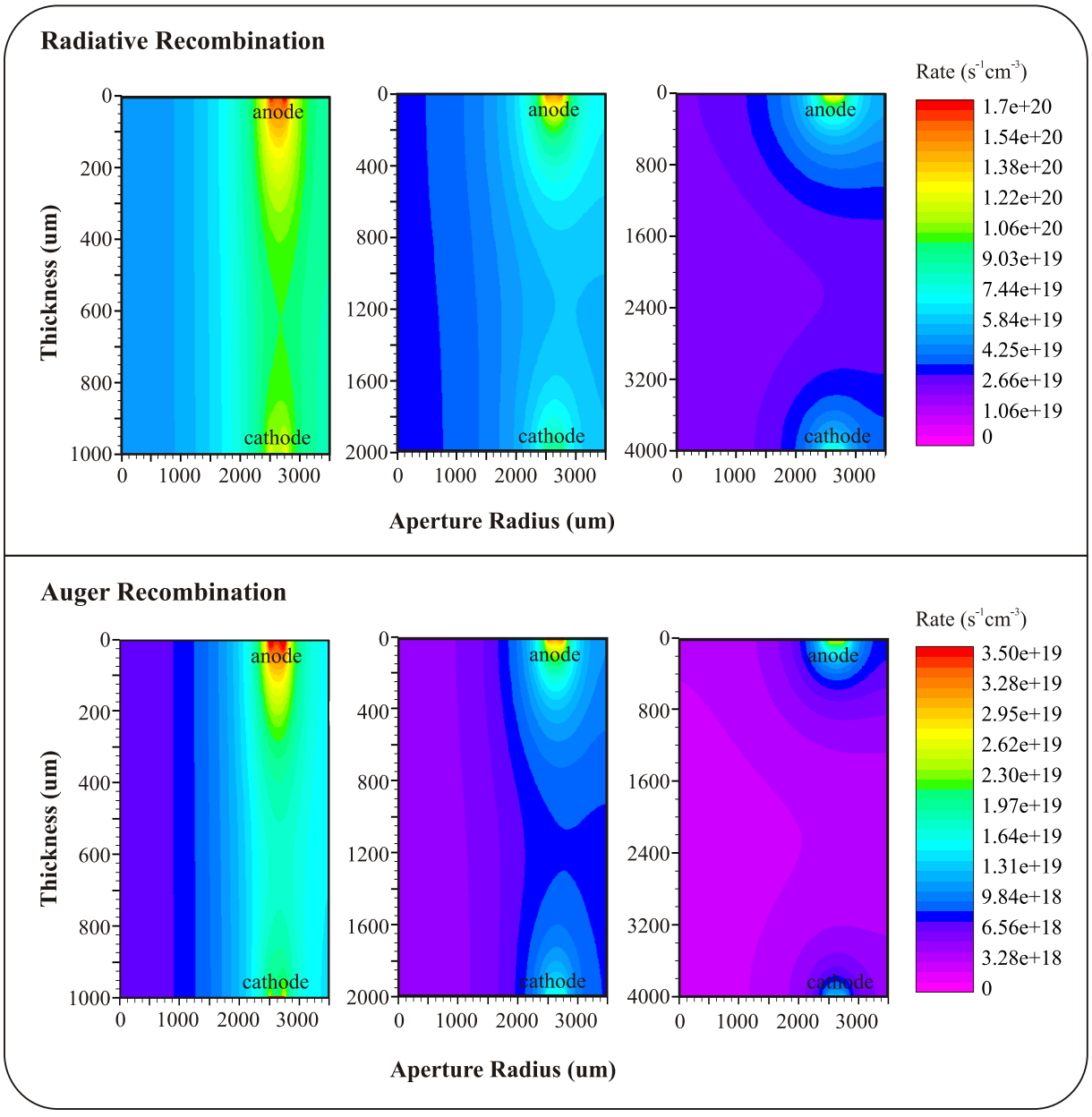


Figure 5.13: The map of carrier recombination as a function of modulator thickness at 2.6 A of current. Radiative recombination dominates within the operating regime of the modulator.

efficient trap centres introduced by copper ions have reduce A_h to below $10^{14}\,\mathrm{cm}^{-2}$, two orders of magnitude lower compared to phosphorus. This equates to less than 10% attenuation at $10.59\,\mu\mathrm{m}$.

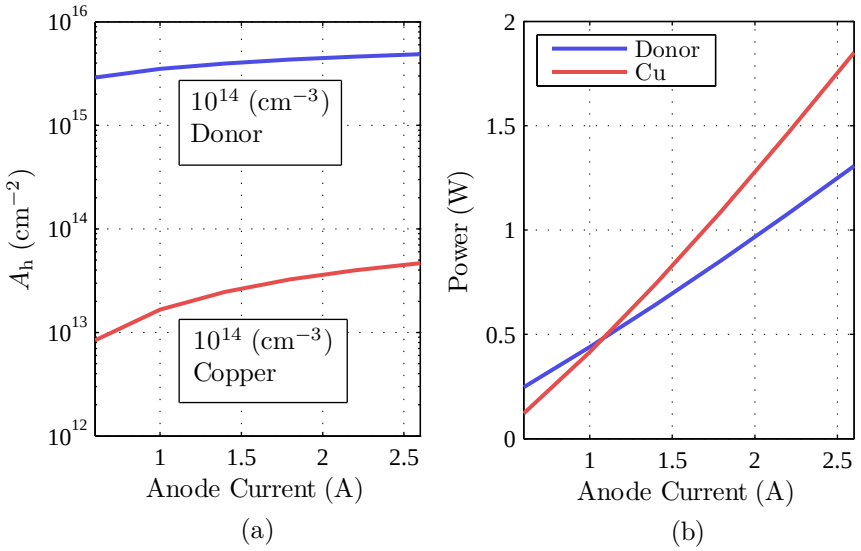


Figure 5.14: The performance of the modulator with different species of impurities added. Phosphorus SRH recombination parameters were used for donors.

Fortunately, high purity crystalline germanium wafers with impurity levels the order of $10^{10}\,\mathrm{cm}^{-3}$ are commercially available (i. e., less than one part per trillion electrically active impurity ions). These were purchased from Umicore[3] for producing the modulators. It is therefore vital that every effort is taken to reduce the diffusion of impurities (especially transition metals) into the substrate during high temperature processing.

Despite the benefits of using intrinsic germanium, it is necessary to dope the modulator with donors to ensure that the absolute transmission remains close to unity at temperatures above 25 °Cwhen its electrically "off". At room temperature, the density of free carriers are low enough to permit close to 99.5% of IR transmission to pass through the modulator (if Fresnel's reflection is ignored). At higher temperatures however, the absolute transmission will start to drop due to the increase in carriers generated thermally. The absorption induced by free carriers can be reduced by lowering the concentration of holes, since its absorption cross section is 15 times greater than electrons [8]. As described by the mass-action law, ($n_i{}^2 = np$ [22]), this can be accomplished by adding donors to the modulator to reduce the concentration of holes

[3] Umicore Electro Optic Materials, Watertorenstraat 3, Olen, 2250, Belgium

(p).

The transmission through a 2 mm thick modulator as a function of temperature and donor concentration is shown in figure 5.15. An N_D of $10^{14}\,cm^{-3}$ would suffice to maintain close to unity transmittance through the modulator for temperatures up to 90 °C.

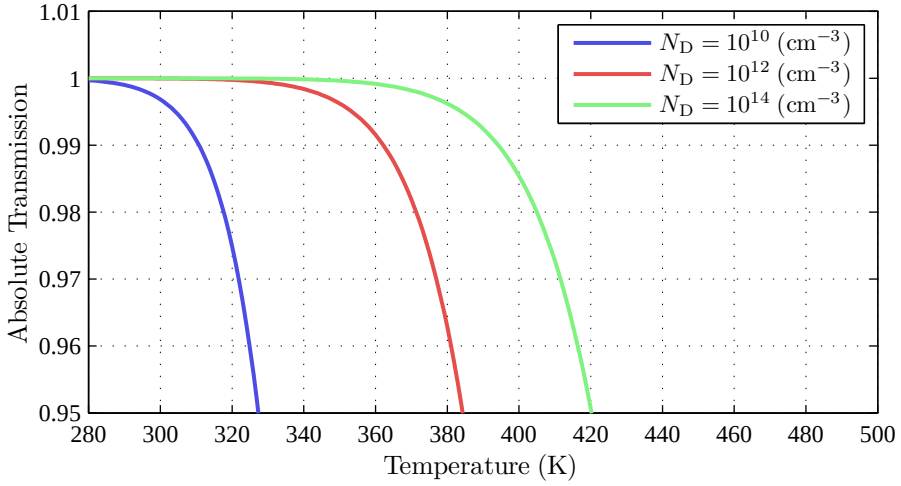


Figure 5.15: The transmission through a 2 mm thick modulator as functions of temperature and donor concentration. A donor density of $10^{14}\,cm^{-3}$ is sufficient to ensure close to unity absolute transmission through the modulator for temperatures up to 65 °C.

The performance of the modulator would not be adversely affected by the addition of a small amount of donors *provided* that shallow dopants are used. This is illustrated in figure 5.14, where a depth of modulation of 92% can be attained (at 2.6 A) using an unoptimized modulator structure.

5.4.4 Wafer Thickness

Having reviewed the adverse effects of recombination mechanisms on the performance of the modulator, results from exploring different techniques to reduce the effectiveness of carrier recombination are presented next.

As depicted in figure 5.13, Auger and radiative recombination becomes considerable during modulator electrically "on" state (i. e., when large numbers of excess carriers are injected into the modulator). Because the rate for each mechanism depends on the density of excess carriers (figure 4.4), the obvious method to reduce their effectiveness is by reducing the density of electrons and holes. The performance of the modulator would not be affected so long as the integral of carrier density (A_h) through the entire depth

of the modulator is unchanged (figure 5.7). Both these requirements can be satisfied with the use of thicker modulators, as results from simulation shows (figure 5.16a).

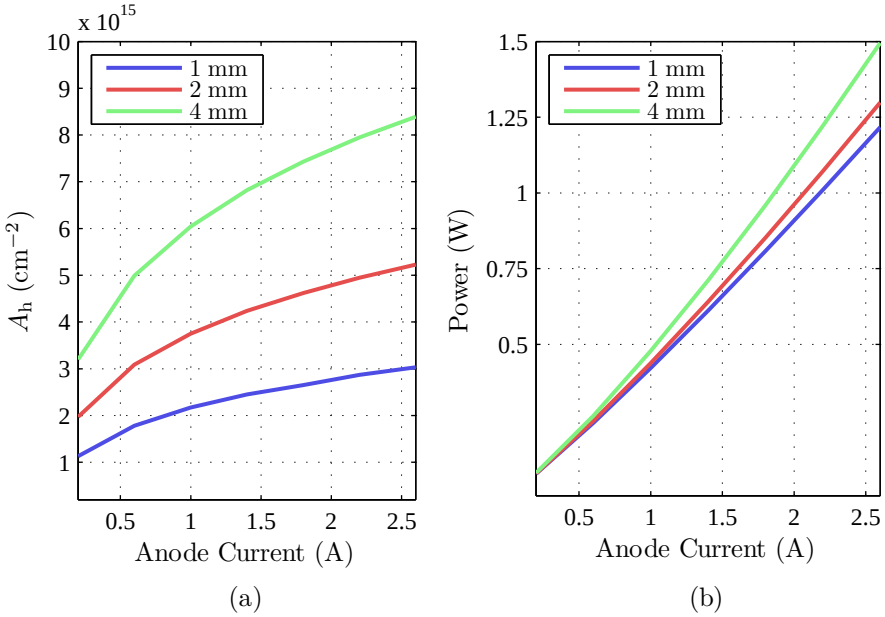


Figure 5.16: The increase in A_h brought about by the reduction in Auger and Radiative recombination.

There is however a maximum limit to the thickness of the modulator. Based on the analysis of carrier dynamics in one-dimension (section 4.2.1), this limit is in the region of twice the diffusion length ($2L_d$). Using wafers thicker than $2L_d$ will result in reduced current modulation, as more carriers would have recombined between the electrodes before they could drift out of the modulator. This in turn lowers the gradient of carrier concentration and ultimately, the depth of modulation. Despite this, when implemented within the bounds of $2L_d$, thickness optimization offers—by far—the biggest improvement on the performance of the modulator.

5.4.5 Reduction in Surface Recombination

Besides bulk recombination, carriers can recombine at the surface via surface recombination (SR). Its rate, being dependent on the density of dangling bonds present at the surface, can sometimes rival that of bulk recombination (although it is odd to compare between the rate of bulk and surface recombination, the equation for calculating the rate of SR is actually the same as the one used for calculating SRH recombination [19]).

Various techniques for reducing the effects of surface recombination (SR) such as

band bending [9, 18], chemical surface passivation [13], and acid etching [14] are available. Chemical wet etching—being the simplest, most accessible and effective method for reducing SR—along with surface band bending have been utilized for reducing SR in our project. The latter was achieved by doping the surface with acceptors to produce pn junction structures (the wafer contains small traces of n-type impurities). The severity of different SR rates on the performance of the modulator are shown in figure 5.17, while the effectiveness of using surface doping to deflect carriers from reaching the surface are summarized in table 5.2.

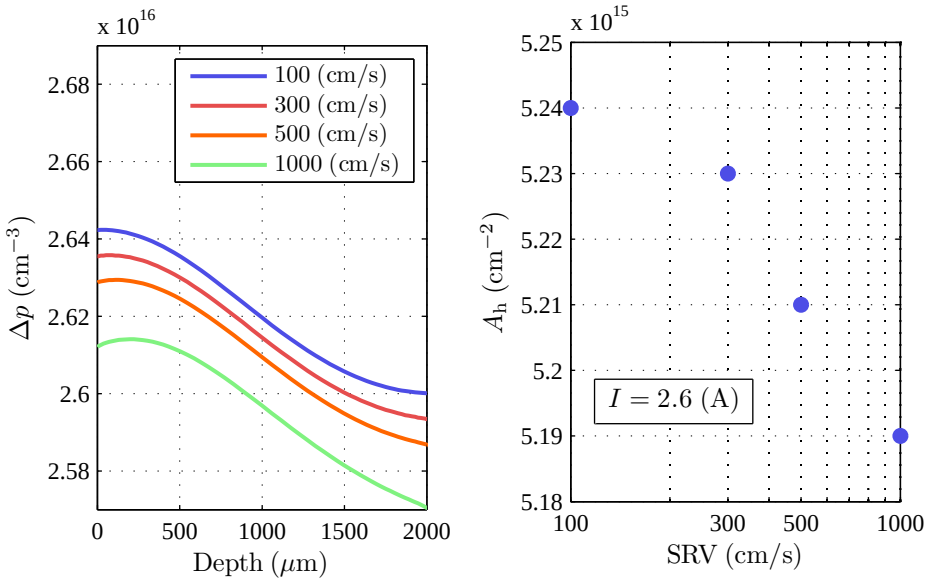


Figure 5.17: (a) The concentration of excess holes through the modulator (in the $\hat{z}$ direction) as a function of surface recombination velocity at 2.6 A, and (b) the corresponding hole area concentration.

Surface recombination velocity (SRV—see section 4.1.6) as low as 10 cm/s can be obtained by etching wafers with acids [14]. Unfortunately, such a low value cannot be maintained for long when samples are exposed to ambient; instead, it raises asymptotically with time. Data from Ponpon showed that the rate of increase of SRV with time is also dependent on the type of acids used [14]. For wafers that are less than a millimetre thick, the deleterious effects of SR can be very severe, where up to a factor of 10 reduction in the effective carrier lifetime is possible (figure 4.7). Beyond two millimetres however, the effects of SR is reduced considerably. As can be seen in figure 5.17, the concentration of holes in the default modulator structure is hardly

affected even at SRV of 1000 cm/s. (The profile of carriers seen, being sampled through the middle of the aperture, is due to the shift in the peak drift current density towards the anode. This is caused by the difference in mobility between electrons and holes.)

Although SRV values from 100 to 300 cm/s can be attained using acid etching, the solutions used can be highly corrosive. Two of the most effective mixtures include HF:H_2O_2 (1:1) and HF:HNO_3 (10:1) [5, 14], which could easily remove a few hundred nanometers of germanium per minute and strip off thin implanted regions on the modulator [26]. (Other types of the mixtures are available, but are not as effective at lowering SRV [5].) Because germanium reacts easily with ambient (even with water), acid etching is only effective when carried out at the end of the modulator fabrication stage, since wafers are repeatedly heated to 120 °C, exposed to UV, and submerged in solvents, photoresist developer and strippers, and metal etchants. Without damaging the layers that form the modulator (thin SiO_2 dielectric, aluminium and antimony electrodes and the 'n' and 'p' doped layers), the only viable choice of solution was to use diluted hydrofluoric acid composed of 10:1 ratio of H_2O to HF. Although this mixture produces acceptable values of SRV between 200 to 300 cm/s, the surface is less stable and pales in comparison with sub 100 cm/s SRV that can be achieved with solutions containing H_2O_2 and HNO_3. Therefore, in efforts to further reduce the effects of SR, surface band bending is employed to compliment acid etching.

Unlike surface etching which removes deleterious states from the surface, thereby reducing SR, band bending deters carriers from reaching the surface instead. Experimental evidence for using pn junction induced surface band bending has been published by Linross [12]. By implanting the surfaces of n-type silicon wafers with acceptors to form $p^+nn^-np^+$ structures, the diffusion of carriers to the surface is impeded by the potential barrier formed in the space charge region.

Backed by *initial* simulation results that predicted an increase in carrier lifetime for modulators with surfaces doped with acceptors, coupled with experimental evidence from Linross supporting the use of surface pn junction structures for reducing SR [12], all bar one of our modulators have been fabricated with heavily doped surfaces. Furthermore, because ATLAS predicted an even bigger improvement for 'n' doped surfaces (as apposed to 'p' doped surfaces to form pn junction structures needed for bend bending—which was very odd), the entire surface area on the anode side was doped with acceptors, while donors were used on the opposite surface. By doing so, the requirement to have highly doped layers below the electrodes, and the need to

produce structures less affected by SR are met simultaneously. The number of steps required to produce the modulator is reduced as a result.

Aperture Surface Implantation

Peak Dopant Concentration (cm^{-3})	n-type ($A_\mathrm{h}\ \mathrm{cm}^{-2}$)	p-type ($A_\mathrm{h}\ \mathrm{cm}^{-2}$)
10^{15}	5.22×10^{15}	5.22×10^{15}
10^{17}	4.91×10^{15}	4.94×10^{15}
10^{19}	1.48×10^{15}	4.44×10^{15}

Table 5.2: The effects of different peak surface dopant density on A_h at 2.6 A.

This decision was ill-chosen as results from subsequent simulations showed that the opposite was true (table 5.2). As the density of dopants at the surface increases, A_h decreases. This was due to the incorrect implementation of SRH recombination in early simulation models[4]. As a consequence, the lifetime of carriers in doped regions were severely overestimated, giving them sufficient time to scatter away from the surface instead of succumbing to SRH recombination.

It must be noted that in experiments conducted by Linross, doped silicon wafers with an average thickness of 500 μm were used. Such thin samples are more susceptible to SR. Thus the use of pn junction structures to deflect carriers is feasible for reducing SR, provided that SRH recombination beyond the space charge region is not significant. But for thick intrinsic germanium where the effects of SR are weak—even negligible if the density of carriers far exceed the density of surface traps, dopant inclusion at the surface will only add to the rate of SRH recombination, leading to reduced modulator performance (table 5.2). Also, because the build-up of electric field in the space charge region is proportional to the doping density [22], band bending is far more effective in doped wafers—as opposed to intrinsic wafers which only contain traces of dopants. Nevertheless, the use of n-type wafers (with donor density up to $10^{14}\,\mathrm{cm}^{-3}$) for future device fabrication (section 5.4.3) will make this a feasible method for suppressing SR.

[4]Instructions given in the ATLAS manual for implementing SRH recombination were ambiguous at best. We had to refer to articles cited in the manual to correctly invoke the model [16].

5.4.6 Optimized modulator Performance

Having discussed different structural alterations that can be made to augment modulator performance, these are then incorporated into the simulation model to produce an optimized structure. Improvements obtained are shown in figure 5.18. For comparison, the performance of structures with pessimistic parameters are also included, with differences between each model listed in Table 5.3.

Simulation Parameter Specifications

Parameter	Optimistic	Pessimistic	Default	Units
Electrode Width	10	250	250	μm
Peak N_A and N_D	5×10^{19}	1×10^{19}	1×10^{19}	cm^{-3}
C_n and C_p	10^{-31}	10^{-30}	10^{-31}	cm^6/s
SRV	250	500	300	cm/s

Table 5.3: Differences in parameters between optimistic and pessimistic models.

As can be seen in figure 5.18, the worst performing structure is capable of achieving 90% depth of modulation at 2.6 A. When optimized, the optically "off" state transmission can be improved by a factor of five. Increasing the thickness from 2 to 4 mm produced even greater results, where transmittance is reduce by an order of magnitude (for optimized structures). Clearly, the use of thicker wafers is very effective in lessening the effects of Auger and radiative recombination. In fact, the design of our modulator makes it necessary that thick wafers be used to achieve good modulation without excess power consumption. The drawback however is the reduction in modulation frequency, as drift and diffusion currents take longer to reach steady state. This is reflected in figure 5.19, where the decay time for carriers in a 4 mm thick sample is several times longer than the duration needed to reach steady state at 2.6 A. Even so, the slow rate of carrier decay in samples as thick as 4 mm does not present a cause for concern, as applications that benefit from the use of the modulator (e. g., systems using pyroelectric detectors, built-in thermal calibrators for imaging systems using photon detectors) require no more than 100 Hz to operate.

We did however study the usefulness of more complex bias voltage waveforms to bolster the operating frequency of the modulator. Improvements obtained are presented in the next section.

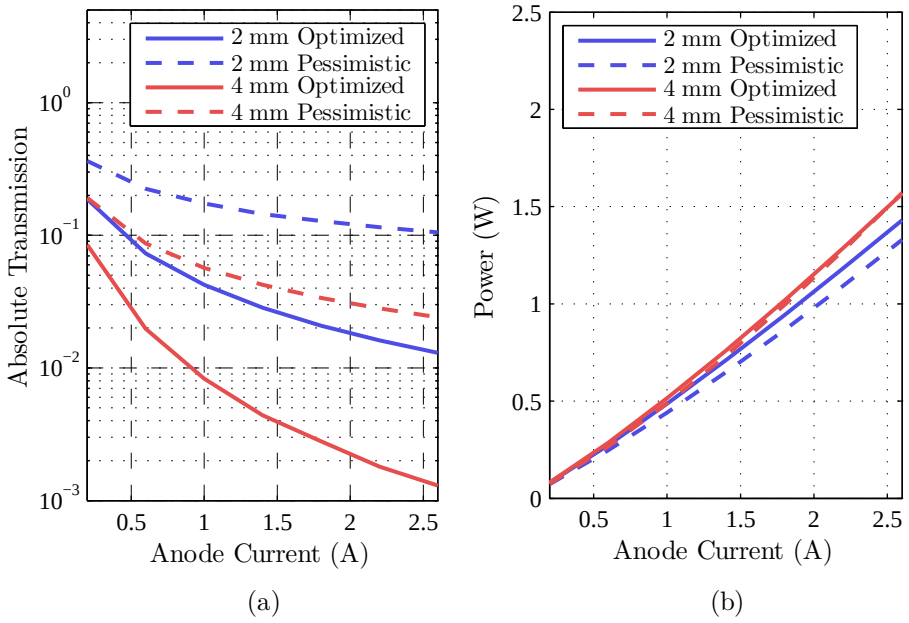


Figure 5.18: Comparison between models with optimistic and pessimistic parameters

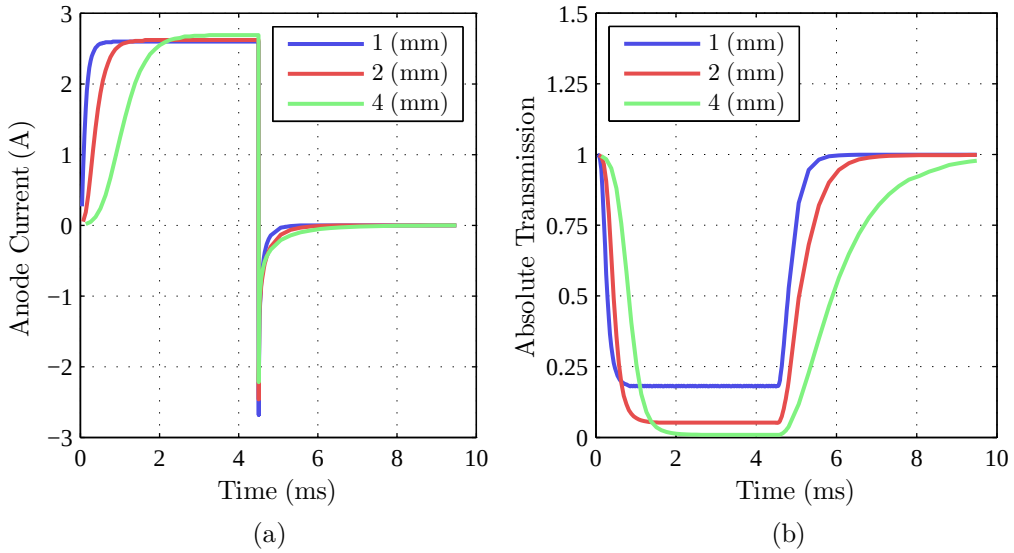


Figure 5.19: The time needed for modulators with different thicknesses to reach steady state.

5.5 Device Optimization: Pulse Width Modulation

In efforts to increase device operation frequency and lower power dissipation, pulse width modulation was used. The modified waveform is shown in figure 5.20, and it can be divided into four parts, from (a) to (d). The function of each part is described as follows: (a) Instead of driving the modulator at a constant voltage, an overshoot is deliberately applied to rapidly inject carriers into the modulator; (b) the voltage required to maintain the required steady state current through the modulator; (c) momentary reduction in bias voltage to reduce power dissipation; and (d) the application of reverse bias to reduce the recovery time of the modulator (i. e., the time for excess carriers to decay to intrinsic values). For comparison, the default bias voltage is shown alongside in red. Because thick structures have been shown to achieve greater depth of modulation at the expense of lower operating frequencies (figure 5.19), pulse width modulation was used on 4 mm thick modulators, as they stand to benefit the most from enhancement in operating frequency.

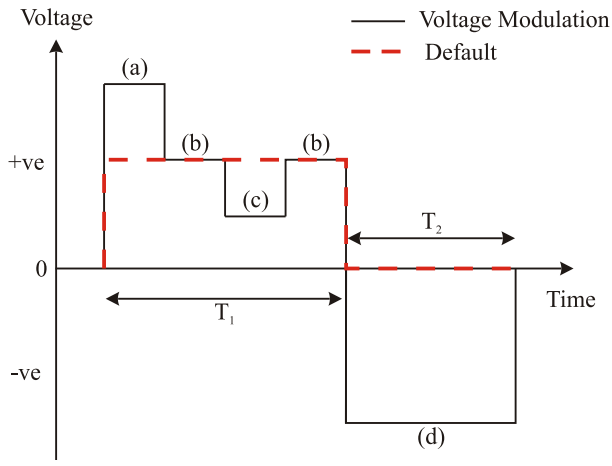


Figure 5.20: The modulated voltage waveform to enhance modulator performance.

5.5.1 Voltage Overshoot

The modulator works by flooding the intrinsic 'i' layer with a sufficient amount of carriers to induce infra red attenuation (IR) between $8 - 14\,\mu$m. These carriers originate from drift currents through the modulator during forward bias (figure 5.6), since the diffusion of carriers is caused by the gradient in carrier concentration. Therefore, to achieve good depth of modulation, the required current flow through the modulator

must first be established. This process can be accelerated by applying a large voltage overshoot to quickly inject carriers into the modulator. The duration and the rise and fall time of the overshoot must be kept short or the exponential rise in current will result in severe power dissipation, as can be seen in figure 5.21.

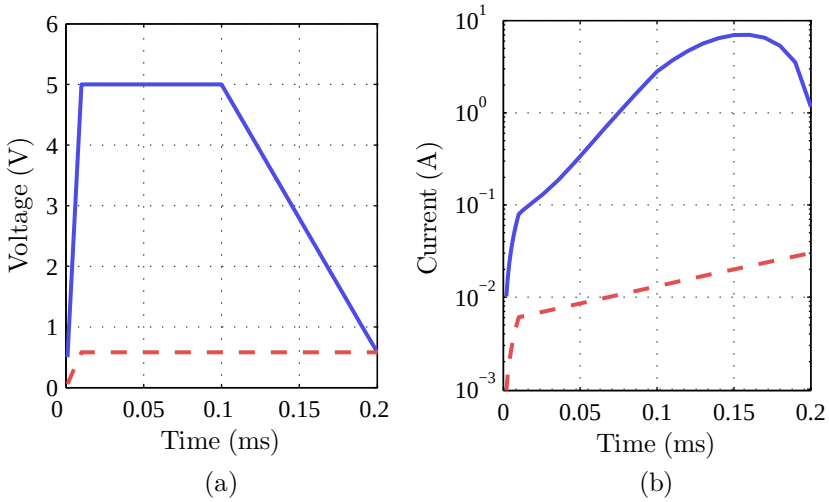


Figure 5.21: The current and voltage waveforms of a 4 mm thick modulator. The default forward voltage signal used in simulation is shown in dotted red lines, while the application of voltage overshoot is shown in blue. The non-linear rise in current will quickly exceed the required steady state current unless the overshoot is ramped down fast enough.

With appropriate pulse duration applied, the intended steady state current of 2.6 A (the current to achieve 95% modulation at 10.59 μm) can be reached much faster. This is shown in figure 5.22. As the current flow through the modulator is increased more rapidly, so too is the rate of carrier diffusion.

So long as the maximum current limit is not exceeded, several voltage pulses can be applied consecutively. Square pulses with fixed amplitude but variable frequencies were used for this purpose. However, the non-linear current-voltage characteristics of the modulator made it extremely difficult to implement the train of pulses with the right amplitude and frequency without causing the current to oscillate violently. With the forward current frequently hurtling above 10 A within a few microseconds, it was down to trial and error to alter the amplitude, duration and separation of each pulse. This proved to be a tedious and time consuming process, in which no suitable waveforms were found within the limited time of the project. Instead of using several voltage pulses, the best results were obtained using a pulse which decays in manner shown in figure 5.23. This waveform was found by trial and error. As can be seen, the intended

magnitude of current flow was kept high compared to the much simpler single pulse waveform shown in figure 5.22.

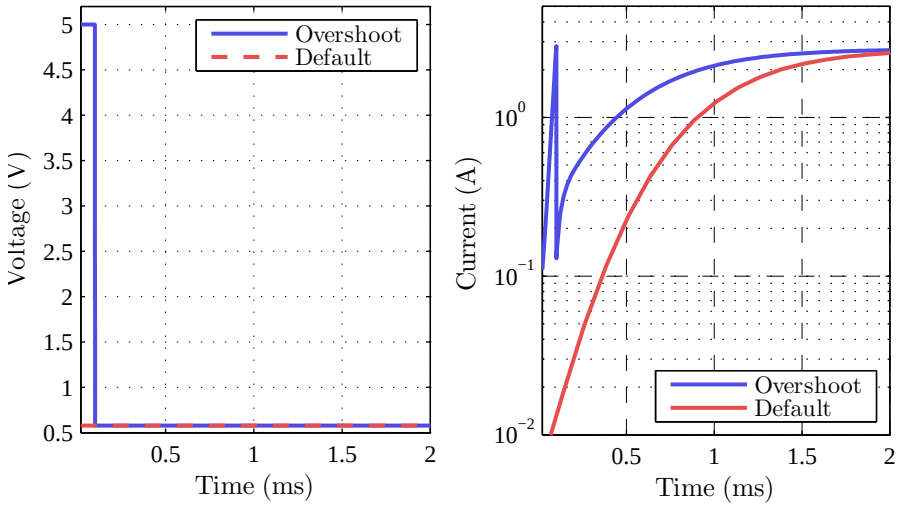


Figure 5.22: The increase in current flow through a 4 mm thick modulator with voltage overshoot applied (the leading edge of the overshoot pulse is shown)

5.5.2 Momentary Reduction in Forward Voltage

As carriers are injected into the modulator during forward bias, they form drift currents that flow through the electrodes. Whilst some of the carriers will diffuse to flood the modulator, the majority are swept away without contributing to device performance. If the modulator is used in low frequency applications or in dc mode, a lot of power will be wasted. A possible solution to lessen this wastage is by momentarily reducing the forward voltage once equilibrium is established. However, as in the case of voltage overshoot, the non-linear current-voltage (IV) characteristics of the modulator made it hard for this method to be implemented properly. When the voltage is reduced, the current also drops rapidly. This in turn disrupted the diffusion of carriers, causing inhomogeneous absorption across the diode. Due to time constraints, this method to reduce power dissipation was not successfully implemented. However, there are reasons to believe that this technique will work, although it is hard to estimate its effectiveness. If the effective lifetime of excess carriers is known (around 0.5 ms for intrinsic germanium. See figure 4.6), and if the current through the modulator can be increased fairly rapidly (a fraction of a millisecond. See figure 5.24), the flow of current

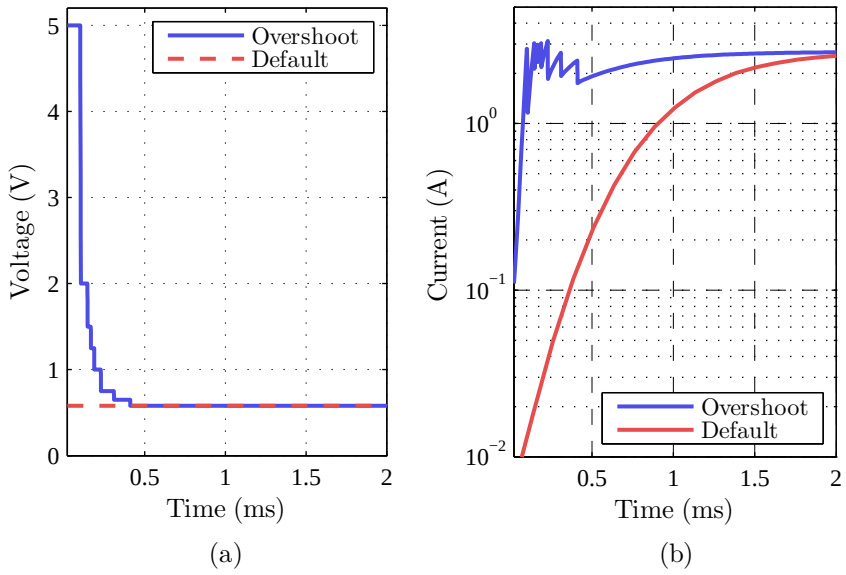


Figure 5.23: Modified overshoot pulse to maintain the required steady state current through the modulator (the leading edge of the overshoot pulse is shown).

through the modulator can be reduced momentarily—so long as the combined duration at which the voltage is in the low state and the time needed for current to be be raised to steady state values does not exceed the effective lifetime of carriers.

5.5.3 Reverse Biased Diode

The time needed for the modulator to fully recover depends on the recombination rate of carriers—the very thing we have strived to minimize in order to achieve good depth of modulation. As a consequence, the operating frequency of the modulator is reduced due to long carrier decay lifetimes. Although the recovery process can be hastened by increasing the rate of carrier recombination (e. g., adding impurities), doing so will impair the performance of the modulator whereby problems such as inhomogeneous absorption across the aperture and low depth of modulation will ensue. Alternatively, diode recovery can be accelerated by the application of a reverse voltage (V_{rev}) in series with a resistor. This is known as the reverse recovery method, and it is normally used for measuring the lifetime of carriers in diodes [1]. The circuit diagram, current and voltage behaviour of the modulator are shown in figure 5.25.

At time (t) less than 0, a forward current I_{d} flows through the modulator. The concentration of excess carriers in the modulator is assumed to be in equilibrium under the constant flow of I_{d}. In this state, the resistance through the diode is low, and the

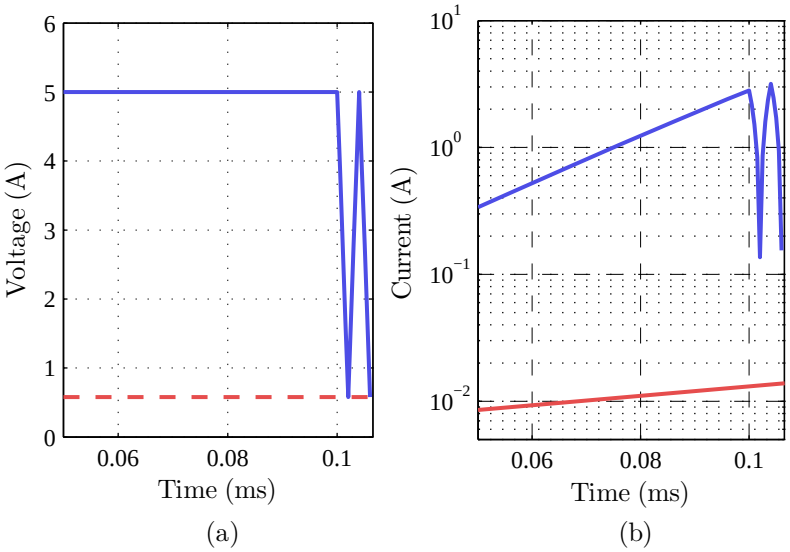


Figure 5.24: The application of voltage overshoot to rapidly inject carriers into the modulator. The oscillatory behaviour of the voltage waveform is mirrored by the flow of current. As can be seen, it only takes a fraction of a millisecond for the current to increase from 0.1 to 2 A. The default voltage and current waveform is shown in red.

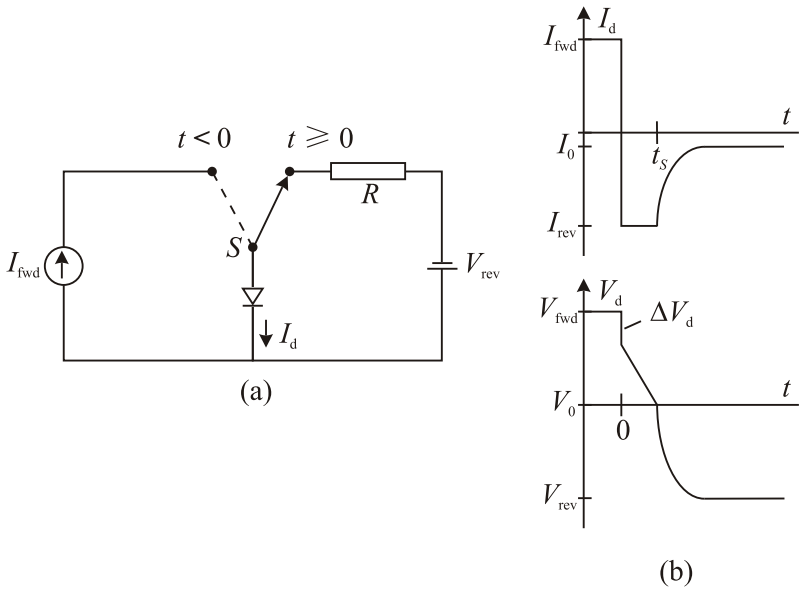


Figure 5.25: (a) Reverse recovery circuit diagram, and (b) current and voltage waveform with abruptly switched current.

voltage across it is represented by V_{d}. Upon reversing the bias voltage at $t = 0$, the current starts to flow in the opposite direction with magnitude given by $I_{\mathrm{rev}} \approx (V_{\mathrm{rev}} - V_{\mathrm{fwd}}/R)$. When this happens, a small drop in the potential across the modulator (ΔV_{d}) occurs due to ohmic voltage drop across the device [6]. Even with the current flowing in the opposite direction, the voltage across the modulator remains positive because it is proportional to the logarithm of carrier density. The large but momentary value of I_{rev} is maintained by the sweeping of excess carries out of the modulator. As more carriers are depleted, space charge regions (SCR) surrounding the electrodes start to form. Carriers that are not swept away by the applied electric field will recombine instead. This is true for carriers that are far away from the depletion region not affected by the electric field. As more carriers are swept away, I_{rev} starts to fall at time t_{s}. This coincides with the edges of the space charge region being almost depleted of excess carriers. As time progresses, the V_{d} approaches the reverse bias voltage V_{rev}, and the current approaches the leakage current I_0. The magnitude of I_0 is determined by the rate of carrier generation in regions depleted of carriers as discussed in section 4.1.1. The recovery process of the modulator is illustrated in figure 5.26.

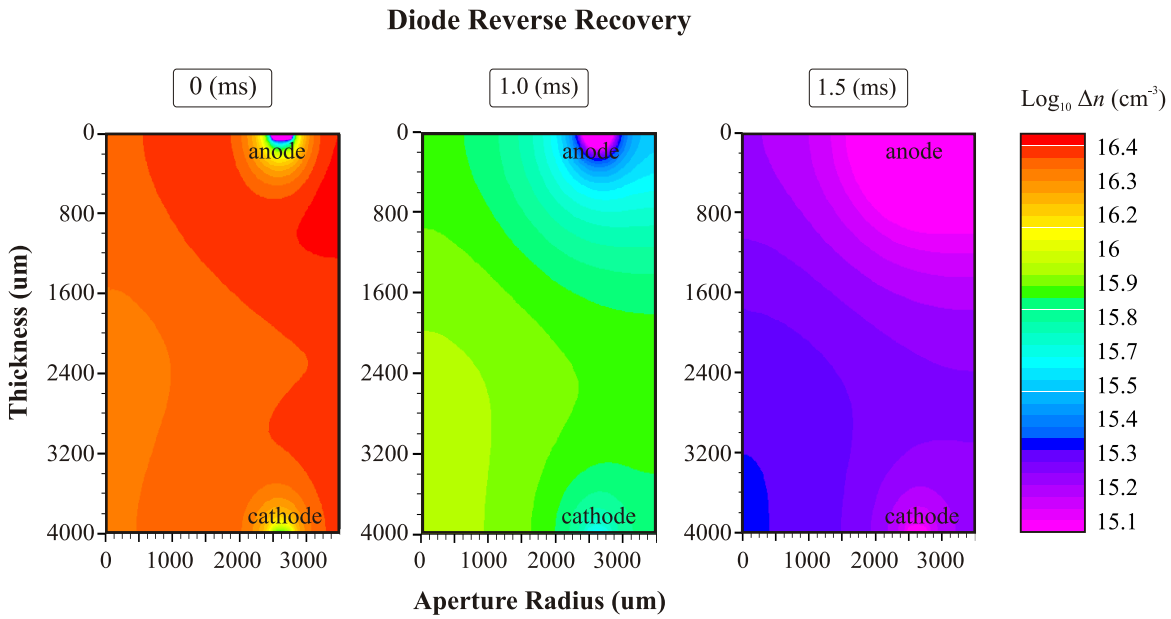


Figure 5.26: The change in carrier concentration at V_{rev} of 10 V. As carriers are swept out from the modulator, regions close to the electrodes become depleted of carriers.

Simulation of the use of reverse bias to aid recovery was performed on the default modulator structure with different reverse voltages. The results are shown in

figure 5.27. For transmittance below 0.6, the effect of reverse voltage is hardly noticeable. Beyond 0.8 however, the recovery times can be cut short of up to a few hundred μs. The impact of reverse recovery is much more significant in thicker modulators as figure 5.28 shows. A decrease by up to 0.5 ms is obtained at transmittance of 0.9 when compared between 0 and -10 V of applied voltage. This equates to $\approx 8\%$ reduction in recovery time, if the fall time is defined as the time needed for the transmittance to rise to 0.9.

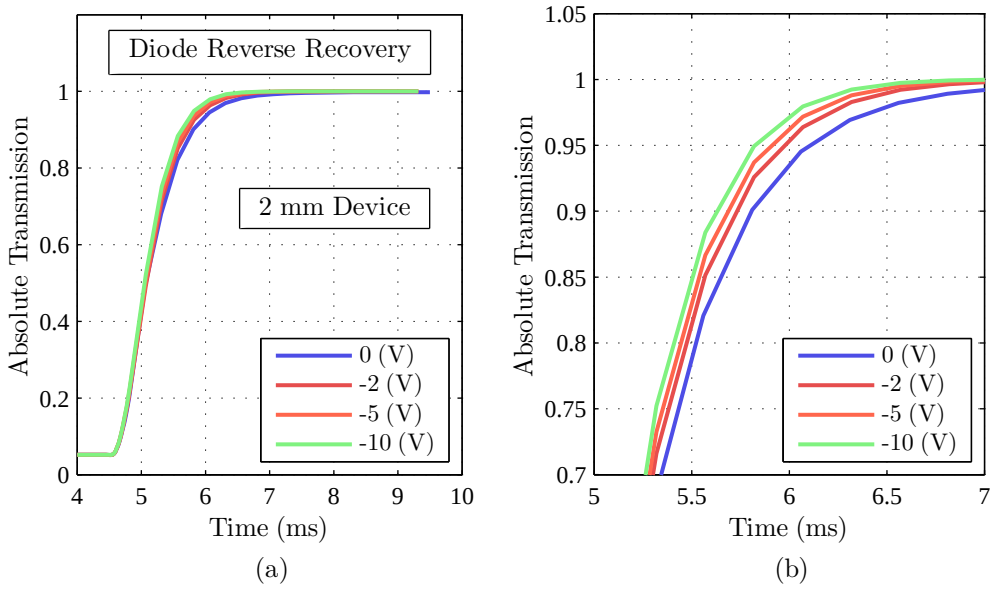


Figure 5.27: Recovery time for the default modulator structure under different reverse bias voltage. The steady state current used was 2.6 A.

Reverse voltage recovery is therefore a viable method to decrease the "turn off" time of the modulator although its effect is modest. By reversing the gradient in carrier concentration through the formation of space charge regions around the electrodes, the outward diffusion of carriers are enhanced, leading to improved recovery time.

5.5.4 Conclusion

The general capabilities of ATLAS along with the steps involved in setting up a simulation model were presented. The electrical behaviour of the default modulator structure was predicted using ATLAS as being capable of producing $\approx 95\%$ depth of modulation at 10.59 μm without dissipating more than 2 W of power. Factors that influence the performance of the modulator such as carrier diffusion and recombination were investigated. The depth of modulation can be improved if the rate of carrier diffusion is

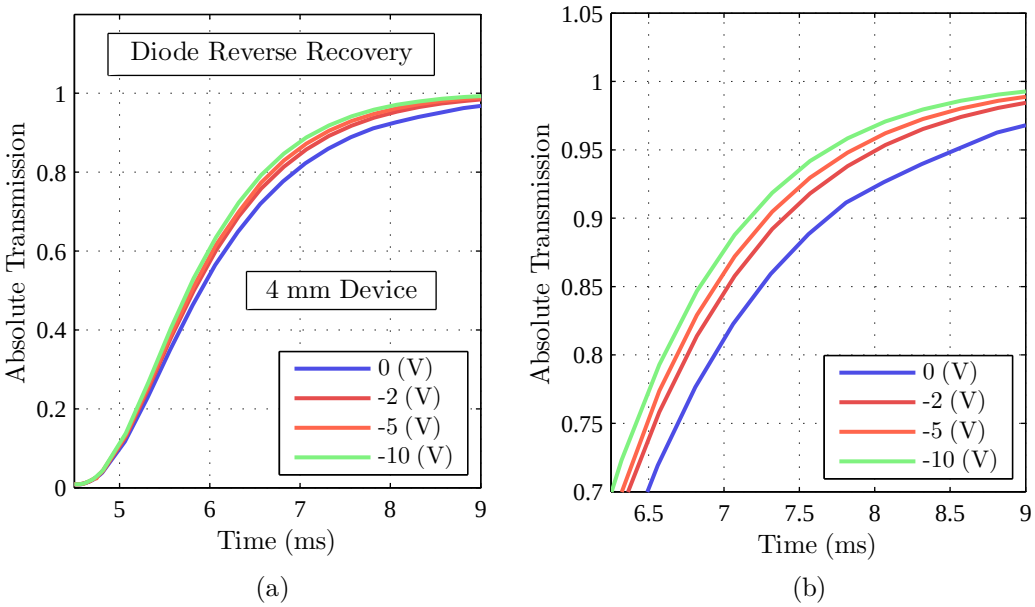


Figure 5.28: Recovery time for a 4 mm thick modulator under different reverse bias voltage. The steady state current used was 2.6 A.

increased. This was achieved through the use of narrower electrodes to increase the current density through the modulator. Methods to reduce recombination mechanisms such as the use of thicker wafers to lessen the effects of Auger and radiative recombination, acid etching and band bending to mitigate surface recombination, and using wafers with extremely low concentration of electrically active impurities to reduce SRH recombination were explored in detail (the wafer can be doped with *shallow* donors up to a concentration of $10^{14}\,\mathrm{cm}^{-3}$ to ensure good transmission through the modulator during optical 'on' state). The optimized parameters were incorporated into the default structure to augment its performance. At 2.6 A, up to an order magnitude reduction in transmittance is obtained when compared with the default structure. The operating frequency of the modulator was increased through the use of forward voltage overshoot and reverse voltage on switch-off. As a result, the rise time for current through a 4 mm modulator was reduced by $\approx 50\%$, while the recovery time was shortened from 7.5 to 7 ms. Though the rise time can be reduced by half, the optimized recovery time is several times longer than the rise time; thus only small improvements to the operating frequency can be achieved using pulse width modulation.

Bibliography

[1] F. Berz. Step recovery of p-i-n diodes. *Solid State Electronics*, 22(11):927–932, 1979.

[2] Y. L. Chao, S. Prussin, and J. C. S. Woo. Preamorphization implantation-assisted boron activation in bulk germanium and germanium-on-insulator. *Applied Physics Letters*, 87(142102):1–2, 2005.

[3] C. O. Chui, K. Gopalakrishnan, P. B. Griffin, J. D. Plummer, and K. C. Saraswat. Activation and diffusion studies of implanted n and p dopants in germanium. *Applied Physics Letters*, 83(6):3275–3277, 2003.

[4] C. O. Chui, L. Kulig, J. Morgan, and Wilman Tsai. Germanium n-type shallow junction activation dependences. *Applied Physics Letters*, 87(091909):1–3, 2005.

[5] P. D. Fairley. *Novel Solid State Modulator for the Infrared: The Germanium Chopper*. PhD thesis, University of Southampton, 2000.

[6] S. K. Ghandi. *Semiconductor Power Devices*. John Wiley & Sons, 1977.

[7] D. J. Griffiths. *Introduction to Electrodynamics*. Prentice Hall, New Jersey, 3rd edition, 1999.

[8] C. J. Hutchinson, C. Lewis, J. A. Savage, and A. Pitt. Surface and bulk absorption in germanium at 10.6 μm. *Applied Optics*, 21(8):1490–1495, 1982.

[9] M. Ichimura, M. Hirano, N. Kato, E. Arai, H. Takamatsu, and S. Sumie. Control of surface recombination of si wafers by an external electrode. *Japan Journal of Applied Physics*, 38:L292–L294, 1999.

[10] C. Jacaboni, F. Nava, C. Canali, and G. Ottaviani. Electron drift velocity and diffusivity in germanium. *Physical Review B*, 24(2):1014–1026, 1981.

[11] V. K. Khanna. Carrier lifetimes and recombination-generation mechanisms in semiconductor device physics. *European Journal of Physics*, 25:221–237, 2004.

[12] J. Linnros. Carrier lifetime measurements using free carrier absorption transients. ii. lifetime mapping and effects of surface recombination. *Applied Physics*, 84(1):284–291, 1998.

[13] D. Poelman, P. Clauws, and B. Depuydt. Chemical surface passivation of low resistivity p-type ge wafers for solar cell applications. *Solar Energy Materials and Solar Cells*, 76:167–173, 2003.

[14] J. P. Ponpon. Evolution with time the surface properties of high-purity germanium. *Nuclear Instruments and Methods in Physics Research A*, 457:262–265, 2001.

[15] L. Reggiani, C. Canali, F. Nava, and G. Ottaviani. Hole drift velocity in germanium. *Physical Review B*, 16(6):2781–2791, 1977.

[16] D. J. Roulston, N. D. Arora, and S. G. Chamberlain. Modelling and measurement of minority carrier lifetime versus doping in diffused layers or $n^+ - p$SiliconDevices. *IEEE Transactions on Electron Devices*, 29(2):284–291, 1982.

[17] A. Satta, E. Simoen, T. Clarysse, T. Janssens, A. Benedetti, B. D. Jaeger, M. Meuris, and W. Vandervorst. Diffusion, activation, and recrystallization of boron implanted in preamorphized and crystalline germanium. *Applied Physics Letters*, 87(172109):1–3, 2005.

[18] M. Schofthaler, R. Brendel, G. Langguth, and J. H. Werner. High-quality surface passivation by corona discharged oxides for semiconductor surface characterization. In *Proceedings on the* 1st *World Conference on Photovoltain Energy Conversion*, pages 1509–1512, New York, 1994. IEEE.

[19] D. K. Schroder. *Semiconductor Material and Device Characterization*. John Wiley & Sons, New Jersey, 3rd edition, 2006.

[20] Silvaco International, Santa Clara, CA, USA. *ATLAS User's Manual: Device Simulator Software*, 2004.

[21] Y. S. Suh, M. S. Carroll, R. A. Levy, G. Bisognin, D. D. Salvador, M. A. Sahiner, and C. A. King. Implantation and activation of high concentrations of boron and germanium. *IEEE Transactions on Electron Devices*, 52(11):2416–2421, 2005.

[22] S. M. Sze and K. K. Ng. *Physics of Semiconductor Devices*. John Wiley & Sons, New Jersey, 3rd edition, 2007.

[23] B. Tien and C. Hu. Determination of carrier lifetime from rectifier ramp recovery waveform. *IEEE Electron Device Letters*, 9(10):553–555, 1988.

[24] F. A. Trumbore. Solid solubilities of impurity elements in germanium and silicon. *The Bell System Technical Journal*, 39:205–233, 1959.

[25] S. Uppal and H. N. Rutt. An electrically driven solid state modulator. In 2nd *EMRS DTC Technical Conference*, 2005.

[26] K. R. Williams, K. Gupta, and Matthew Wasilik. Etch rates for micromachining processing—part ii. *Journal of Microelectromechanical Systems*, 12(6):761–778, 2003.

[27] P. V. Zant. *Microchip Fabrication*. McGraw Hill, 5th edition, 2004.

Chapter 6

Modulator Fabrication

The procedure to produce a fully functional modulator using semiconductor wafer processing technology is presented in this chapter. The fabrication method was chosen because it satisfies both the stringent requirement to keep the wafers clean from organic and metallic contaminants, and also the need to produce micron size features on the modulator. Since advances in wafer fabrication have been focussed on technologically important materials such as silicon and gallium arsenide [36], relatively little is known about germanium wafer processing[1]. We therefore had little choice but to process our wafers using many procedures and chemicals meant for processing silicon wafers. This approach was reasonable due to the similarity between silicon and germanium (e. g., same crystalline structure, adjacent to one another in group IV of the periodic table), though it is well known that germanium is chemically more reactive compared to silicon [18]. Nevertheless, many of the processes had to be 'tweaked', optimized or abandoned altogether for producing the modulator. The bulk of the chapter therefore details the problems encountered during fabrication and the solutions devised to overcome them. For this reason, the basic operations involved in wafer fabrication will only be reviewed *very* briefly in the first few sections, since there is already plenty of literature dedicated to them [16, 40]. The chapter ends with a summary of important steps involved in the production of the modulator, including the attachment of copper heat sinks and the method used in forming electrical contacts to the diode.

[1] This is about to change as more literature on germanium wafer processing is published by the microelectronics community to revive the use of germanium in developing future integrated circuit devices [31].

6.1 Basic Wafer Fabrication Operations

There are essentially four basic operations involved in wafer processing: material deposition, patterning, doping and heat treatment [40]. Material deposition is used to add thin layers of material to the wafer surface. Metal, dielectric and photoresist are some of the most common examples of layers deposited on wafer surfaces. Patterning is performed to remove selected portions of materials/layers to obtain the desired geometric shape (also known as windows or patterns). This entire process is also known as lithography, and is used together with the material deposition step to sequentially deposit layered structures onto wafers. The doping step involves incorporating electrically dopants into wafers using a variety of methods including ion implantation and high temperature diffusion. Finally, the heat treatment step involves heating the wafers to temperatures ranging from 250 to 1000 °C for various purposes. These include to repair/recrystallize amorphous regions in wafers [16], to activate dopants incorporated via ion implantation [9], and to reduce contact resistance by "alloying" metal electrodes with the semiconductor [40].

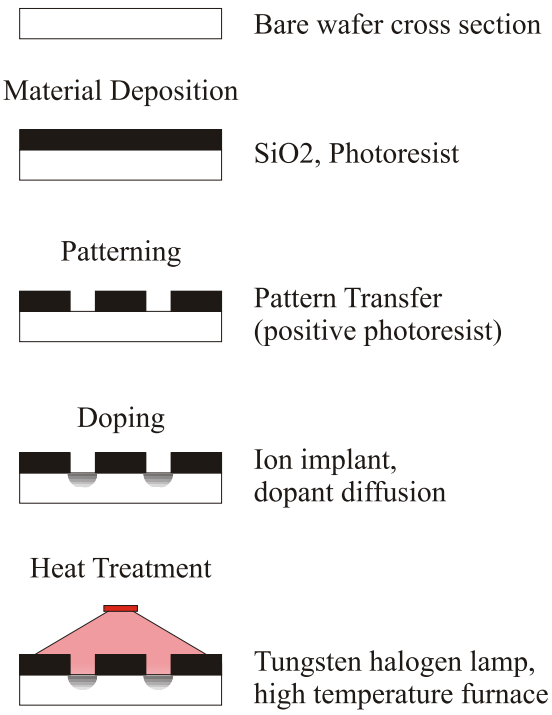


Figure 6.1: The basic semiconductor fabrication steps

6.2 Photolithography

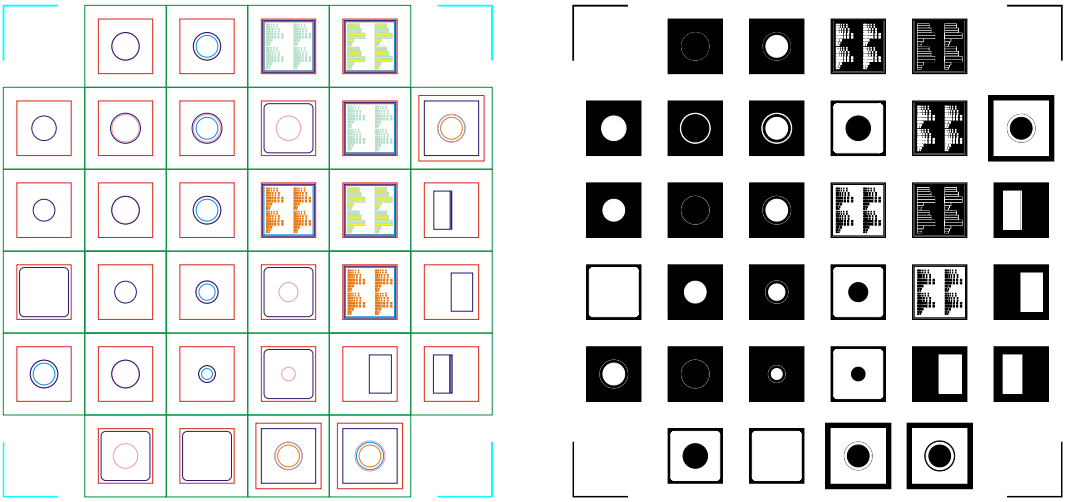
Photolithography is used for transferring patterns defined by a mask to a layer of UV sensitive material (called photoresist) that is deposited onto a wafer surface. The process generally proceed as follows: (a) spin coating photoresist onto the wafer surface, (b) bake the wafer to drive solvents out of photoresist, (c) expose the wafer with a UV source through a patterned mask with the layer of photoresist sandwiched between the wafer and mask. Regions on the photoresist exposed to UV then undergo a chemical reaction which, depending on the type of photoresist, would either render them soluble (positive photoresist) or insoluble (negative photoresist) to developers, and (e) removing specific regions of the photoresist using developers to form intended patterns on the wafer. Details on the use of positive and negative photoresist are given in sections 6.3 and 6.9.1 respectively.

6.2.1 Masks

Masks are usually made from fused silica plates with a thin layer of deposited chrome. Fused silica is used because of its low thermal expansion coefficient, high transmittance between 200 and 400 nm and good mechanical strength [34]. Using electron beam or deep UV lithography in ultra clean processing environments, patterns with feature size < 500 nm can be transferred to the layer of chrome [34]. This is the reason why masks are relatively expensive to make. Our project however does not require such fine structures, and could tolerate larger sized defects since the smallest feature size in the modulator is roughly $50,000$ nm wide! Masks for this project were instead produced by J. D. Phototools[2] using low resolution lithographic technology that combines the use of an image plotter and reduction lenses to transfer patterns onto the mask. A minimum feature size between 4 and 5 μm can be produced. In addition, defects larger than 5 μm on the chrome layer are on average no more than 0.15 per square centimeter, which is sufficient for our purposes. To further lower cost, the masks were made using soda lime glass that has a transmittance value from 85 to 88% at 375-450 nm and a flatness tolerance around 5 μm. Both these specifications would be disastrous for standard submicron CMOS processing, as inhomogeneity in UV exposure and the diffraction of UV light prior reaching the photoresist caused by the large flatness tolerance would result in a very low yield. Again, these do not present a problem to us due to the large

[2] JD Photo-Tools, Meridian Centre, King St, Oldham, OL8 1EZ

feature size used. The design of the mask used for fabricating the modulator is shown in figure 6.2b. It is drawn using CorelDraw[3], a commercial drawing software. Instructions and rules for designing the mask are readily available in standard semiconductor fabrication texts [16, 40].



(a) (b)

Figure 6.2: (a) The CAD design of the mask that contains all the pattern required to produce the modulator and (b) the actual mask image used in fabrication.

6.3 Photoresist

The photoresist (PR) is a UV sensitive compound that is used for transferring patterns to wafers through photolithography. It can be classified as positive or negative depending on the chemical reaction that takes place after UV irradiation, and generally has an absorbance greater than 0.6 below the wavelength of 350 nm [16]. Because photoresists have good degree of resistance towards acids, they are often used as barrier layers for transferring patterns to selected regions of the wafer through the patterning step, which involves either adding or removing materials in regions not covered by PR. Once the patterns are transferred, PR can be stripped off using suitable solvents without harming the patterns.

Positive photoresist (PPR) consist of three components: a photosensitive compound, a base resin and an organic solvent [34]. Exposing PPR to UV will render it

[3]Corel USA, 46430, Fremont Blvd. Fremont, CA 94538,United States of America

soluble to developers; thus regions previously exposed to UV light will be removed after the development stage, completing pattern transfer to the PR.

Negative photoresist (NPR) is a mixture of polymers and a photosensitive compound. It works by converting the absorbed optical energy from UV irradiation into chemical energy to decompose certain compounds in the resist. A post exposure bake below 100 °C will initiate a cross linking reaction in exposed regions to form highly resistant and durable epoxy layer. Pattern transfer is completed after immersing wafers in developers to remove unexposed regions.

Positive photoresist is used in the patterning step because it is easier to process. It can even be stripped off using standard cleanroom solvents such as acetone and isopropanol. Negative photoresist on the other hand requires 3 to 4 additional steps for patterns to be transferred, and is cumbersome to handle since its unexposed state forms a sticky resin when dissolved in acetone. More importantly, once cross linking is initiated, NPR is extremely hard to remove. Manufacturers have suggested using immersion in strong oxidising acids, plasma ash, reactive ion etch and even laser ablation to remove hardened NPR, all of which are not viable for our work. NPR was however used as a dielectric layer to produce our first functional modulator before it was replaced with SiO_2 (section 6.9.1).

6.3.1 Pattern Transfer

Our wafers were sawn into 10 by 10 mm^2 pieces to increase the number of samples available. All fabrication steps are thus performed using them. S1828 positive photoresist produced by MicroChem [4] was used in photolithography, and the guidelines for transferring patterns to this resist are outlined in figure 6.3.

To obtain maximum process reliability, the sample must be cleaned and the surface dehydrated prior to resist coating. This is done by immersion in an acetone filled beaker placed in an ultrasonic cleaner for 15 minutes. The solvent can be heated to 50 °C to enhance the cleaning process. The sample is then rinsed with isopropanol and deionized water to remove any residues. This is followed by a dehydration bake in the oven at 120 °C for 30 minutes. Once dried, adhesion promoters can optionally be applied to enhance resist adhesion (these were used in section 6.9.2). This is normally used if PR is be exposed to acids at the patterning stage. After the baking step, the sample is placed on a chuck that is attached to a vacuum spindle, and PR is applied

[4] MicroChem Corp. 1254 Chestnut Street, Newton, MA 02464

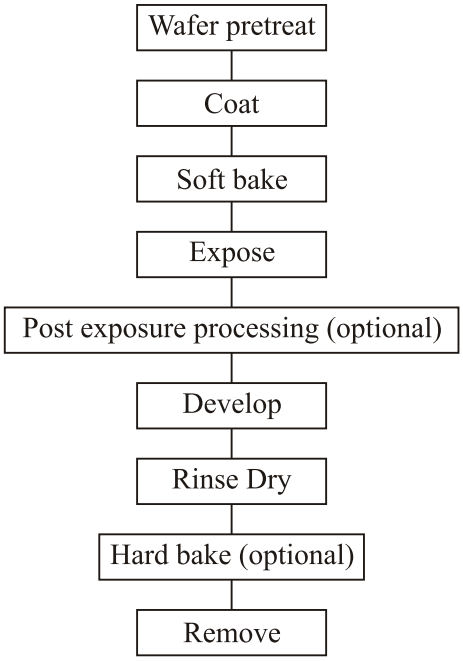


Figure 6.3: Positive photoresist process guidelines

to the sample using a syringe with 0.45 μm particulate filter attached. Although PR is usually applied to the centre of circular wafers until $> 70\%$ of surface area is covered, doing so produced many streaks in our square shaped samples. Injecting droplets of PR at the corners produced better results. Once resist is applied, the chuck is ramped to 5000 rpm within 5 seconds, and left to dwell for 30 seconds. This coats the samples with a uniform layer of film about 2.5 to 2.8 μm thick. Samples are then baked for 40 minutes at 90 °C to remove solvents from the PR. Upon completion, the sample is aligned with respect to the mask in an optical lithography system (Karl Suss MA6 UV mask aligner), and exposed with UV light for 13 seconds at a power density of 14 mW/cm^2. This is followed by a 45 second immersion in developer (ME319 from MicroChem) to dissolve away regions exposed to UV. The pattern transfer process is completed by rinsing the sample with deionized water and blow dried using dry nitrogen. An example of patterns transferred to the PR is shown in figure 6.4.

6.3.2 Overhang Structures

Thick electrodes ($\approx 1\,\mu$m) have to be deposited on the modulator to ensure that good contacts are formed in our device. This is done after the pattern transfer process (see figure 6.4), where the structure of the electrode is already defined on the photoresist.

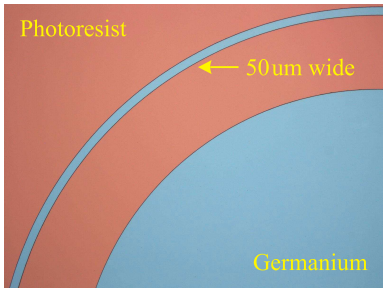


Figure 6.4: The pattern of the annular electrode and the aperture successfully transferred to a layer of photoresist.

A layer of metal is then deposited, and resist stripped off to transfer the pattern from the PR to the sample. This process is called liftoff and it is illustrated in figure 6.5.

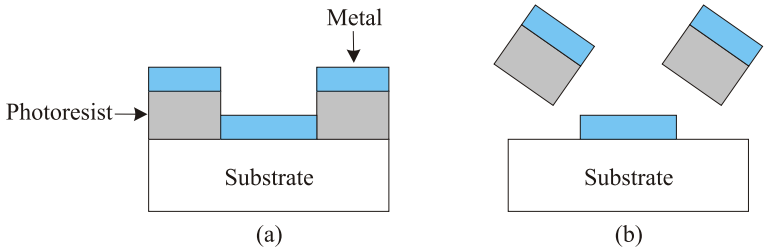


Figure 6.5: Liftoff process for pattern transfer

To prevent the shearing of metal during liftoff (especially for a very thick layer of metal), it is crucial that overhang structures as shown in figure 6.6 are produced to aid the liftoff process. The desired resist structure must be taller than the thickness of the metal layer to eliminate bridging of the metal on top of the resist [40]. In addition, the overhang must also be large enough so that the metal is not deposited on the sidewalls of the photoresist.

Overhang structures can be formed using a single layer of photoresist or from the combination of two dissimilar layers of materials. In the latter, photoresist is spun on a layer of material that can be etched away in a solution that does not erode the photoresist. Once patterns are transferred to the photoresist, the sample is soaked in a solution that isotropically etches the layer under the photoresist (i. e., similar vertical and horizontal etch rate) to form overhang structures over time. In the interest of keeping the number of fabrication steps low, the first option was pursued[5].

[5]Michael Pollard, a summer project student produced the first overhang structure in the course

The steps required to produce overhang structures in a single layer of PR involve: (a) soaking the sample in chlorobenzene for 10 minutes after the UV exposure stage, (b) develop sample in standard developer for 5 minutes, and (c) rinsing sample with deionized water and blow dry using nitrogen. When chlorobenzene penetrates the surface of photoresist, it becomes more resistant to the developer (the developer does etch away unexposed photoresist at much slower rates). Overhangs can thus be created by immersing the sample in the developer that undercuts the bottom region of the photoresist. An example of overhang structures can be seen in figure 6.6b.

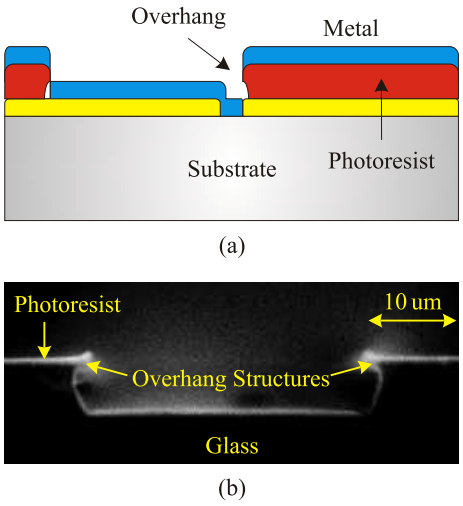


Figure 6.6: (a) Overhang structures used when depositing thick metal layers, and (b) actual overhand structures created using positive photoresist.

6.4 Thin Film Deposition Methods

Several metal and dielectric layers are needed to produce the modulator (figure 6.34). Two methods have been employed for depositing these materials: thermal evaporation (TE) and electron beam evaporation (EB). Both are fairly easy to use and give fast turnaround times. Though EB is the preferred choice of evaporator, dielectric materials such as ZnS (for anti-reflection coating) and toxic antimony (electrical contact) had to be evaporated in dedicated TE chambers to prevent cross contamination of cleanroom equipments. Both methods also offer good control over the rate of evaporation, where rates between 0.5 to 10 nm/s can be easily achieved.

of the project using a single layer of photoresist. The author is indebted to him for all his work in optimizing the liftoff process.

6.4.1 Thermal Evaporation

In thermal evaporation, the material (charge) to be evaporated must first be placed in a boat, or crucible inside a vacuum chamber. Once the chamber is evacuated to pressures less than 3×10^{-6} mBar, a current between $50 - 200$ A is passed through the boat or the basket shaped filament which holds the crucible to resistively heat the charge. Depending on the material, it could either form a melt before evaporating, or it could sublimate. The flow of vapour, unimpeded under high vacuum conditions, then condenses on the surface of the sample to form a layer. To improve thickness uniformity, the sample is mounted on a rotating plate. The TE process is shown in figure 6.7.

Compared to electron beam evaporation, the main advantage of using TE is its gentleness, where the stoichiometry of dielectric compounds can be better preserved (i. e., compound materials such as ZnS less likely to dissociate upon evaporation). Its main disadvantage is its inability to evaporate refractory materials such as titanium. TE has been used to evaporate zinc sulphide (ZnS), magnesium fluoride (MgF_2), tin (Sn), indium (In), and aluminium (Al) in our project. EB was subsequently used for evaporating Al, since metal boats (made from tungsten, molybdenum or tantalum) used in TE systems frequently cracked due to crystallization of alloys formed between molten Al and the metal boats [17]. Al films produced also have a yellowish hue instead of a silver sheen observed in EB evaporated films. (Although alumina coated boats are available for evaporating Al in TE systems, they often require more current than our systems could provide. Smaller boats on the other hand could not hold enough charge to obtain the required thickness.)

6.4.2 Electron Beam Evaporation

In an electron beam evaporation system, materials are evaporated by focussed high energy electrons (typically $\sim kV$). These electrons are given off by a heated filament under high vacuum, which are directed to the source using a magnet. A pool of molten charge is formed in the region irradiated with electrons (usually at the centre). At the same time, the outer edge of the charge remains in the solid state due to water cooling of the crucible. This prevents the molten material from reacting with the crucible, thus reducing the number of contaminants diffusing into the charge. Evaporation begins by increasing the filament current to produce more electrons to heat up the molten material. The rate of evaporation is controlled by the current through the filament.

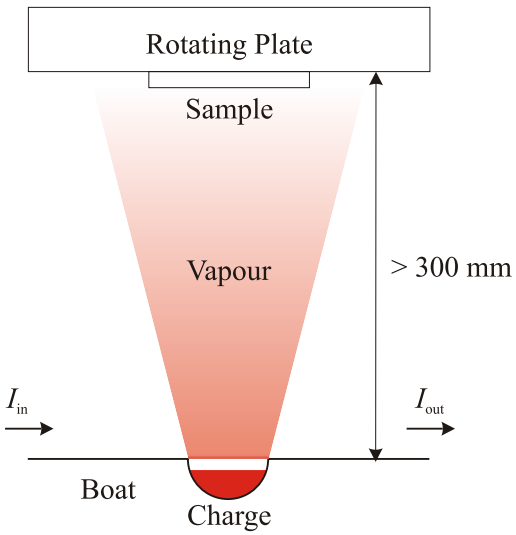


Figure 6.7: The evaporation of source material in a vacuum chamber

Samples are mounted in a rotating plate (not shown) to ensure that materials are evenly deposited.

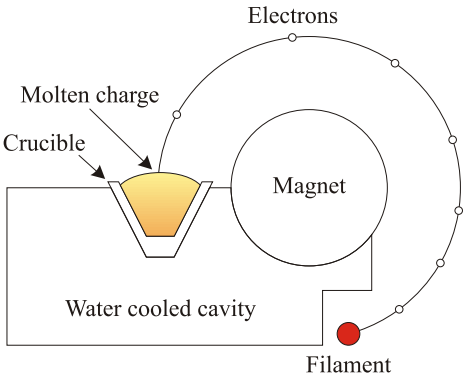


Figure 6.8: The evaporation of source material in an electron beam system.

EB offers several advantages compared to TE. Very thick films can be evaporated because of the large crucibles used. Since only *localized regions are heated*, power dissipation can be kept low (unlike TE where the whole crucible is heated to the melting temperature of the source to initiate evaporation, further aggravating the problem of cross contamination). Finally, refractory materials such as titanium and tantalum can be easily evaporated because the electrons are focussed to a small spot and most of the kinetic energy of electrons is converted into thermal energy upon bombarding the surface of the charge. This however is the reason why compound materials such as ZnS

and MgF_2 are more likely to dissociate, resulting in loss of stoichiometry.

6.4.3 Thickness Monitor

In thermal and electron beam evaporation systems, the rate and thickness of film deposited are observed in-situ using quartz crystal monitoring. The method is based on the principle that the oscillating frequency of a quartz crystal is changed by the mass of a deposited film on its upper face. Electronically measuring this effect allows for a determination of the thickness of a deposited film. Once the density of the evaporated material is entered into the system, the thickness can be measured to a resolution of 1 angstrom and a rate resolution of 0.01 angstrom per second. The deposition rate is very important because it influences the structure and the surface morphology of films deposited [1].

All major processing steps involved in producing the modulator, starting from the bare wafer to the finished product will now be presented sequentially in the remaining part of this chapter.

6.5 Wafer Properties

Intrinsic single crystal germanium wafers were used to fabricate the modulators. Each wafer contains less than $10^{10}\,cm^{-3}$ of electrically active impurity concentration and an average dislocation density less than $3000\,cm^{-2}$. For these reasons, the effective carrier lifetime under low injection densities can be up to tens of milliseconds long (section 4.1.3 and figure 4.6). Though expensive, high purity germanium ingots can be purchased commercially, since they are grown for making gamma-ray detectors [12]. Our wafers were custom prepared by Umicore (>£100 per piece), where 1 and 2 mm thick wafers were sawn from a 50 mm diameter ingot grown using Czochralski pulling technique (CZ) [40]. Although single crystal germanium with much lower impurity content can be produced using the float zone method [40], the diameter of ingots grown are less than 10 mm (CZ grown ingots could reach up to 50 mm) because the ratio of the melt surface tension to the density is too small to support a larger floating zone [12]. The wafers are [1,0,0] orientated, specified because most work on germanium were conducted using this orientation (e. g., ion implantation, solid phase epitaxy [9, 30]). Except for work involving alloying (section 6.8.2), we have no evidence that other crystallographic orientations are better for producing the modulator. After the wafers

are sawn from the ingot, the surfaces are polished and etched with HF:HNO_3 to remove subsurface damage and deleterious states from the surface (section 6.6.4).

6.6 Surface Preparations

The condition of the surface is one of the most important factors that determines the performance of the modulator. The presence of interfacial layers such as organic contaminants, thin film residues formed by exposure to different chemicals, and large particles will affect the quality of contacts formed; subsurface damage and dangling bonds will increase the rate of surface recombination, which in turn reduces the depth of modulation; and the presence of oxides will hamper efforts to dope the surface of germanium using alternative doping methods such as solid phase epitaxy. Additionally, if the surface is contaminated with transition metals (TM) and is later exposed to high temperature processing (e. g., annealing at 600 °C to activate implanted dopants), the lifetime of carriers will be reduced due to the diffusion of impurities into the sample (i. e., increased SRH recombination). Therefore, proper surface preparation procedures must be carried out prior to important fabrication steps (e. g., dielectric layer deposition, electrical contact formation, and dopant incorporation) to ensure that a functional device is produced.

Wet chemical treatment was used for cleaning and etching the samples (i. e., soaking/rinsing in solutions). Although the option of using dry cleaning is available (e. g., reactive ion etching, plasma etching), it is only effective for certain applications such as the removal of organic and metallic contaminants, but is not good at removing large particles [40]. In addition, dry etching is known to introduce defects onto the surface of wafers which will degrade the formation of electrical contacts and worsen the effect of surface recombination [10, 14].

6.6.1 Solvent Cleaning: Removing Organic Contaminants

Samples have to be cleaned several times during the course of modulator fabrication. The standard method for removing organic contaminants in silicon processing is to soak wafers in concentrated nitric acids (>85% HNO_3), or in a mixture of NH_4OH, H_2O_2 and H_2O [16]. The latter is known as RCA-1, and it involves the oxidation and complexing chemistry based on hydrogen peroxide and acids. Another frequently used solution is Piranha, which consist of 50:1 ratio of H_2SO_4 and H_2O_2. It is a strong oxidiser that can

remove most organic matter, but it must be maintained at 120 °C to be effective [39]. Though these solutions are not harmful to silicon, the surface of germanium is easily corroded by them due to reaction between germanium and H_2O_2, which would result in mild "orange peel" surface morphology (figure 6.10) to the formation of etch pits similar to ones seen in figure 6.9.

We therefore had to use standard cleanroom solvents (acetone and isopropanol) to clean our samples. These work well provided that ultrasonic cleaning is used. The drawbacks of using this method are: (a) samples have to be cleaned for a long time in the ultrasonic bath to remove stubborn organic residues (e. g., hardened photoresist, partially developed negative resist, and UV cured epoxy used when wafers are sawn to pieces), and (b) prolonged exposure to ultrasonic agitation will peel off layers deposited on samples. For this reason, only 7 of the 21 pieces of germanium that were used to fabricate the modulator survived with every layer intact to form a complete device. (The samples are cut into 10 by 10 mm^2 pieces with a mixture of thicknesses between 1 and 2 mm.)

6.6.2 Acid Etch: Removing Metallic Contaminants

To ensure good carrier lifetimes, it is vital that metallic contaminants are removed from the samples prior to high temperatures processing, or they will diffuse into the modulator to form deep level traps (see sections 4.1.3 and 5.4.3). The removal of metallic contaminants is carried out prior to the deposition of the SiO_2 layer on intrinsic samples (see figure 6.34), and before dopant activation annealing, where the surfaces of samples have been implanted with dopants at depths no deeper than 200 nm. Again, the standard method of removing metallic contaminants involves the use of H_2O_2; in this case, it is mixed with deionized water and hydrochloric acid at ratios that works best between 75 to 85 °C (e. g., between 6:1:1 and 8:2:1 ratio of H_2O, H_2O_2 and HCl [40]). This solution can be used prior to dielectric deposition, where the samples are just square pieces of intrinsic germanium, but not before dopant activation annealing, since it will rapidly etch away the doped layer at the surface (roughly 120 nm/s etch rate for intrinsic germanium and possibly faster for doped surfaces). With no alternative mixtures available (in the literature), we have resorted to using 100:1:1 ratio of H_2O:H_2O_2:HF to etch the surface prior to dopant activation. Though very dilute, recent publications have shown that it is as effective as the more concentrated version used in standard silicon wafer cleaning processing [40].

6.6.3 Native Oxide Stripping

The removal of native oxide from germanium samples (GeO_2) is a crucial step to produce ohmic contacts to the modulator (section 3.4.2). This procedure is easy to carry out because GeO_2 is not a stable compound and even dissolves in water [39]. We have used HF acid to remove the native oxide from the surface of germanium, because Ponpon reported that immersion in this solution will retard the growth of oxide [25]. However, the 1:1 ratio of HF:H_2O recommended by Ponpon resulted in the formation of etch pits on *highly doped* surfaces after 5 minutes of soaking. This is shown in figure 6.9a. HF acid is not known to attack crystalline germanium [39], thus the pits observed may be caused by the inclusion of dopants at the surface. (At time of writing, we could not find any published information on the effects of acids on highly doped germanium surfaces.) To restrict the formation of etch pits, the ratio of HF to water was reduced, and we found that a 10:1 ratio of H_2O:HF did not cause corrosion at the surface even after 10 minutes of immersion.

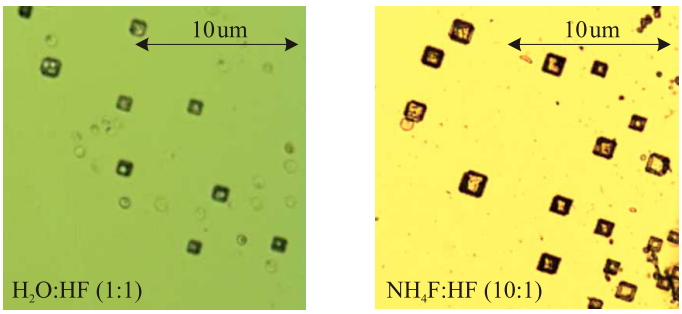


Figure 6.9: Etch pits at wafer surfaces caused 5 minute immersion in (a) strong HF acid, and (b) 10:1 ratio of buffered oxide etch (BOE)

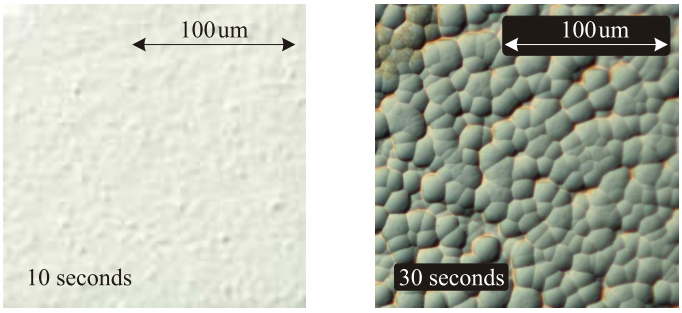
The use of 10:1 H_2O:HF however presented a new problem in device processing: the patterning of SiO_2 layer (step 4 in figure 6.34). At this concentration, it would take around 10 to 15 minutes to etch away 300 nm of SiO_2. Such a long soaking duration caused the photoresist layer to peel off due to undercutting. One of the methods that was explored to overcome this problem is to use buffered oxide etch solutions, which are blends of 49% HF and 40% ammonium fluoride (NH_4F) in various predetermined ratios. BOE gives better SiO_2 etch uniformity because the addition of NH_4F to HF controls the pH value and replenishes the depletion of fluoride ions. More importantly, the *average* etch rate for a 10:1 ratio of NH_4F:HF BOE is almost twice that of 10:1

H_2O:HF. For this reason, it was hoped that the low concentration of HF used in 10:1 BOE mixture would simultaneously reduce the etch time for SiO_2 (to prevent the peeling of photoresist) and preserve the condition of the surface. Unfortunately, the surface of doped germanium was not compatible with BOE, as larger pits were formed after immersing samples in this mixture for 5 minutes (figure 6.9b). In the end, dilute acid (10:1 H_2O:HF) was used to remove native oxide from samples. (The method used to successfully transfer patterns to SiO_2 is detailed in section 6.9.2.)

6.6.4 Surface Passivation

As stressed in section 5.4.3, the surface of the modulator must be properly terminated to reduce the deleterious effects of surface recombination (SR). The easiest and most effective method to achieve this is through wet chemical etching, where surface recombination velocity values (SRV) less than 300 cm/s can be routinely produced. A plethora of mixtures have been reported to reduce SR in germanium [14, 25]. However, the most effective at lowering SR is 1:1 ratio of $H_2O_2 : H_2O$. Not only does it produce sub 100 cm/s SRV values that is stable up to 7 days [25], it is also does not leave a layer of thin film on the substrate (some nitric acid based etchants and other formulation of H_2O_2 do form films on etched surfaces [14]). However, this mixture will rapidly etch the surface of germanium, removing an average thickness of 3500 nm within 30 seconds at 25 °C. Fairley has reported that low SRV values does not necessarily correlate with the complete removal of subsurface damages, as prolonged exposure to etchants often degrade the quality of the surface instead, increasing SRV. Therefore, only a thin layer of germanium needs to be removed for the purpose of reducing SRV. The morphology of surfaces exposed to 1:1 $H_2O_2 : H_2O$ for 10 and 30 seconds are shown in figure 6.10. As can be seen, prolonged exposure to the etchant will result in the roughening of the surface which takes the form of "orange peel" contours at the surface. This could be indicative of the increase in SRV as pointed out by Fairley [14].

As with RCA-2, this solution cannot be used after dopants are implanted into the samples. Therefore, it is performed only once, right after the RCA-2 clean step (section 6.6.2). However, we are fortunate in that the dilute HF acid used for removing native oxide and for patterning SiO_2 is also reasonably effective at lowering SRV ($<$300 cm/s) [25]. Therefore, the modulator is dipped in dilute HF acid prior to the deposition of anti-reflection coating, which doubles as a protective layer to preserve the low SRV values.



(a) (b)

Figure 6.10: The morphology of germanium surfaces etch with 1:1 $H_2O_2 : H_2O$ for (a) 10 seconds and (b) 30 seconds.

6.7 Surface Doping: Ion Implantation

Highly doped 'p' and 'n' type layers are needed below the anode and the cathode of the modulator to achieve high carrier injection efficiency and low contact resistance (sections 4.2.1 and 3.4.3). These layers can be made thin (even down to 20 nm) to reduce the overall IR attenuation induced by excess carriers, which are contributed by dopants (especially acceptors, since the absorption cross section for holes is 15 times greater than that of electrons at 10.59 μm). However, the peak doping level in these layers must be in the order of $10^{19}\,cm^{-3}$, as it is the peak dopant concentration that strongly affects the electrical performance of modulator (section 5.4.2).

The best method to achieve high doping density in a shallow layer is through ion implantation, which introduces energetic, charged particles into a substrate. Ions can be implanted at energies between 1 keV and 1 MeV, resulting in ion distribution with average depths ranging from 10 nm to 10 μm [34]. After implantation, the sample has to be annealed at relatively low temperatures (between 400 to 600 °C for germanium) to restore the crystalline state of the lattice and to position the ions in substitutional sites. The latter is known as the activation process, where dopants sitting in substitutional sites may, depending on their valence state, either contribute electrons, or holes, or neutrally integrate into the host's crystal structure. Ion implantation also offers good reproducibility and lower processing temperature compared to the standard diffusion and alloying process (section 6.8.2) [40]. It was therefore the preferred method for producing highly doped layers on the modulator, though we have also produced functional devices using solid phase epitaxial regrowth (section 6.8.3). In the sections that follow, the criteria for selecting the suitable dopants for the modulator along with

the corresponding implant parameters used, and the resulting doping profiles obtained are detailed.

6.7.1 Choice of Dopants

Various 'n' and 'p' type dopants have been successfully implanted in germanium to high concentrations [9, 28]. These include boron and gallium for acceptors; and phosphorus, arsenic and antimony for donors. Four criteria are used for choosing the most suitable species of dopants for implantation: (a) high levels of electrical activation, (b) low implantation damage to the lattice, (c) low activation temperature, and (d) availability.

For acceptors, two viable options are available: boron and gallium. Both are capable of reaching high levels of electrically active doping density in germanium (between 10^{19} to $10^{20}\,\mathrm{cm}^{-3}$); can be activate at temperatures below 450 °C; and permit the crystal structure to be restored to a good crystallinity even at very high levels of implant dosage (note that the dose is expressed as the number of ions implanted into $1\,\mathrm{cm}^2$ of sample surface area) [7, 9, 28, 33]. Boron was chosen because of its widespread availability in many wafer fabrication centres.

It is worthwhile mentioning that the attainable peak electrically active concentration for boron in germanium via ion implantation is much higher compared to the published solid solubility limit (table 6.1). This has been attributed to the effects of implant damage on the substitutional behaviour of boron in germanium [7]. In fact, boron doping level can be increased by up to an order of magnitude from 10^{19} to $10^{20}\,\mathrm{cm}^{-3}$ by pre-amorphising the surface of germanium via ion bombardment [33]. Doing so retards the channelling of ions (i. e., the unimpeded movement of ions along the major crystallographic direction of the host lattice), confining most of the dopants within the amorphous region. A subsequent annealing step then positions the boron ions in substitutional sites of the host crystal to levels not possible using standard diffusion methods [38]. Remarkably, the density of defects in the recrystallized region after annealing is relatively low ($<10^7\,\mathrm{cm}^-2$; a Czochralski grown crystal ingot may have defect densities up to $10^4\,\mathrm{cm}^-2$ [40]). Nevertheless, the resulting morphology of the surface can be very rough. If the surface is not amorphised, boron ions will tend to form clusters in the lattice, which do not contribute to the total concentration of holes in germanium. Boron also has a peculiar behaviour: once implanted, it has negligible diffusion coefficient even if the germanium sample is heated close to its melting point [38]. This is vastly different from the behaviour of donors such as phosphorus and arsenic,

which diffuses rapidly during annealing. The immobility of boron ions is beneficial to the fabrication of the modulator because the peak dopant density can be maintained even after annealing.

Dopant Element	Type	Solid Solubility (atom/cm^3)
B	Acceptor	5.5×10^{18}
Ga	Acceptor	4.9×10^{20}
P	Donor	2.0×10^{20}
As	Donor	8.1×10^{19}
Sb	Donor	1.2×10^{19}

Table 6.1: Solid solubility limit for common dopants in germanium [37]

For donors, the choice between phosphorus, arsenic and antimony is straight forward. The attainable peak electrically active concentration of phosphorus ($5 \times 10^{19}\,\mathrm{cm}^{-3}$) is roughly 1.5 times higher compared to arsenic, and an order of magnitude greater than antimony [9]. Phosphorus implantation also causes less damage to the lattice compared to As and Sb which, at the required implant dose and beyond, causes irreversible damage to the lattice and even forms abnormal voids at the surface [28]. Phosphorus was therefore used to form highly doped 'n' layers on the modulator. The drawbacks of using phosphorus (as with As and Sb as well), is its rapid diffusion during annealing. Although a temperature of 500 °C is sufficient to regrow the implanted region and simultaneously achieve very high level of donor activation without changing the doping profile of phosphorus, it is insufficient to remove all implant related defects [9]. Temperatures between 600 and 700 °C are necessary for this purpose. At these temperatures, donor ions receive enough thermal energy to rapidly diffuse in the host crystal, which leads to the formation of the characteristic concentration-dependent box-shape doping profiles (fig 6.12). For this reason, phosphorus must be implanted at doses higher than required to retain the level of electrically active donor concentration after annealing. Fortunately, doing so using phosphorus will not adversely affect the crystallinity of the regrown region [28].

6.7.2 Implant Parameters

To estimate the ion dose and energy needed for implanting boron and phosphorus ions into germanium to the required level of electrically active concentration ($10^{19}\,\mathrm{cm}^{-3}$;

see section 5.4.2), the following Gaussian distribution is used:

$$n(x) = \frac{S}{\sqrt{2\pi}\sigma_{\mathrm{p}}} \exp\left[-\frac{(x - R_{\mathrm{p}})^2}{2\sigma_{\mathrm{p}}^2}\right]. \tag{6.1}$$

S is the ion dose (atom/cm^2), R_{p} the projected range, which is the distance from the surface to the depth at which ions are located in the substrate, σ_{p} the projected straggle which is the statistical fluctuations in the projected range, and x is the distance from the surface.

In deciding the implantation depth and ion dose, two important aspects must be considered: the total infrared attenuation during modulator electrical off state and the minimum depth (i. e., the projected range) of the implanted layer at which the modulator can be processed comfortably. Shallow layers risk being etched away during processing; while higher doses are needed to accommodate longer projected ranges to ensure a high level of peak dopant concentration. The increase in the latter will result in the exponential drop in the transmission when the modulator is electrically "off". There is therefore a trade off between the choice of the implant dose and depth.

Because the aim of the project was to build a proof of concept prototype of an electrically operated modulator, ion dose for boron as high as $5 \times 10^{14}\,\mathrm{cm}^{-2}$ was specified in the second batch of wafers to be processed (our last batch of wafers). This allows the acceptor ions to be buried at depth of 100 nm—deep enough for the modulators to be fabricated without damaging the implanted layer. Assuming that 20% of the total number of boron is activated after annealing (figure 3.6), this results in $\approx 5.5\%$ reduction in the absolute transmission at 10.59 μm. The highest dose used for phosphorus is $8 \times 10^{14}\,\mathrm{cm}^{-2}$, resulting in <1% attenuation in absolute transmission (remember that the absorption cross section for electrons is 15 times smaller compared to holes). For future device processing, lower S can be specified to reduce IR attenuation. The drawback however is the resulting thin implantation depth (<20 nm), which may be easily etched away.

The implant parameters for two batches of wafers processed during the course of this project are listed in table 6.2. As can be seen, the implant doses for both species of dopants were increased in the second batch to accommodate longer projected ranges. In addition, BF_2^+ was used in place of B^+ to amorphise the surface during implantation. This was done to enhance the incorporation of boron in substitutional sites during annealing. Shallow projected ranges for phosphorus were specified to account for the rapid diffusion of phosphorus during annealing. (Annealing the sample

Batch No.	Dopant	S (cm^{-2})	Ion Energy (keV)	R_p (nm)	σ_p (nm)
1	B^+	1×10^{14}	20	50	40
	P^+	5×10^{14}	20	20	15
2	BF_2^+	5×10^{14}	106	100	69
	P^+	8×10^{14}	36	31	21

Table 6.2: Implantation parameters for boron and phosphorus ions

at 600 °C for 10 seconds could broaden the doping profile of phosphorus by up to 100 nm; see figure 6.12). The implant energy for both species to reach the designated R_p and the resulting σ_p were determined using SRIM, a group of programs which calculate the stopping range of ions into matter using a quantum mechanical treatment of ion-atom collisions [41].

6.7.3 Sample Implant Preparations

Ion implantation for the first batch of wafers was carried out for us by Innos, while the remaining batch of wafers by staff at the Surrey ion beam centre. Prior to implantation, wafers were cleaned using dilute RCA-2 mixture followed by a dip in H_2O_2:HF (1:1) for 10 seconds to reduce the SRV (section 6.6). Room temperature implantation for both batches of wafers were conducted using parameters listed in table 6.2. The ion beam was tilted by 7° to the substrate surface normal to minimize channelling [8].

6.7.4 Dopant Activation: Rapid Thermal Annealing

After implantation, wafers must be annealed to restore the crystallinity of damaged regions and to activate the dopants. This can be done using conventional furnace annealing or via rapid thermal annealing (RTA). The latter is capable of very high temperature ramp rates (>50 °C/s) and a has fast cool down time, allowing samples to be annealed quickly. In the interest of maintaining the profile of implanted dopants and to reduce the diffusion of impurity ions into the wafer, RTA was used.

In the RTA system utilized (done commercially for us in a wafer fabrication facility at Eindhoven), implanted germanium wafers are placed on intrinsic silicon wafers held inside a argon purged chamber made from quartz. Samples were radiatively heated using arc lamps, and the temperature of the wafers measured using a *non-contact* optical pyrometer that detects infrared energy radiated from the heated sample. (Al-

though optical pyrometers are frequently used for temperature measurement since they do not suffer the drawbacks of wafer contact, they must be properly calibrated for a specific semiconductor material. This is caused by the difficulty in determining the exact emissivity of the wafer at different temperatures and surface conditions, and due to stray radiation competing with signal from the wafer (especially at temperatures below 350 °C) since the radiation energy emitted by the wafer is proportional to T^4 as described by the Stefan-Boltzmann law). For these reasons, the actual temperature at which the wafers are annealed cannot be ascertained accurately, though the errors should be small above 350 °C. Our samples have a temperature tolerance of ±50 °C at 600 °C, which is well above the 3% error in temperature control for standard RTA systems [5]. (Too low a temperature will result in low percentage of dopant activation while high temperatures will aggravate the effects of phosphorus diffusion.)

To protect the wafer surface from reacting with the ambient at high temperatures, a 300 nm thick SiO_2 capping layer is deposited on both sides of each wafer. Although thinner layers can be used (<45 nm), a thicker layer was deposited so it can double as a dielectric layer needed to fabricate the modulator (fig 6.34). This reduces the number of processing steps.

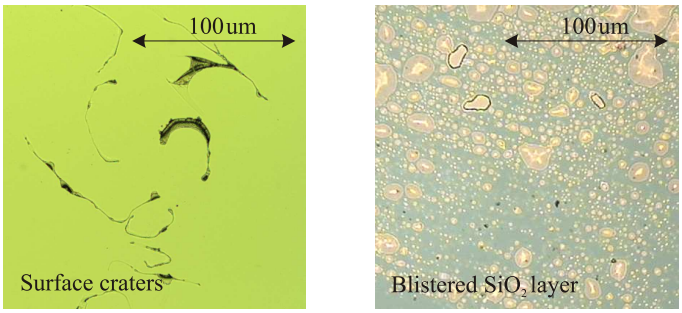


Figure 6.11: (a) ruptured wafer surface due to improper temperature ramp rate, and (b) the blistering of SiO_2 layer after annealing

Wafers are annealed at 600 °C for 10 seconds. A linear ramp profile with a slope of 25 °C /s was initially used (the slope is 30% lower compared to standard RTA recipe for silicon wafers [40]). Two major problems were encountered using this ramp profile: (a) surface rupturing, and (b) the blistering of SiO_2 layer. These are depicted in figure 6.11. The first problem was solved by reducing the ramp rate from 25 °C /s to 5 °C /s and by allowing the wafers to dwell for 20 seconds at set temperatures of 300

and 450 °C. In the second problem, blisters on the SiO_2 layer only occur on the surface of the wafer that rest on the silicon holder. This suggests that they are caused by wafer arching due to the gradient in temperature within the wafer. Reducing the ramp rates down to 5 °C /s did not solve this problem. We even tried to improve the adhesion of the SiO_2 layer to the wafers by depositing the dielectric layer at a higher temperature (from 250 °C to 300 °C). Nothing worked. In the end, SiO_2 layers on both sides of each wafer were removed after annealing by immersing in dilute HF, and fresh layers of SiO_2 redeposited to form the required dielectric layer to produce the modulator.

6.7.5 Dopant Profiling

Using equation 6.1, the distribution of boron and phosphorus for the first batch of wafers is shown in figure 6.12. Included also is the actual dopant profile measured using secondary ion mass spectrometry (SIMS) *after* annealing, which resulted in the broadening of dopant profiles[6]. The profiles were measured for us by Cascade scientific, a company that specializes in surface and materials analysis.

To measure the profile of *electrically active* dopants, spreading resistance profilometry (SRP) is used (SIMS measures the total dopant content in a sample while SRP only measures the fraction of dopants that are electrically active, that is, dopants with dissimilar valence electrons to the host crystal that contribute a specific type of carrier to the system). In this method, the "spreading" resistance between two probes is scanned along the bevelled surface of a sample (the resistance is termed in such a way because the current from one of the probes spreads out radially from the tip). The conversion from spreading resistance data to a carrier density profile and subsequently the doping profile is a complex task that include data smoothing to reduce measurement noise, a deconvolution algorithm, and the use of the correct model for the contact [29]. Again, SRP measurements were carried out by Cascade scientific. The electrically active dopant profiles for the first batch of implanted samples are shown in figure 6.13.

Two striking differences can be seen in figure 6.13 when results between SIMS and SRP are compared: (1) the electrically active dopant density is much higher than

[6]SIMS is a destructive method of analyzing materials whereby the surface of the sample is slowly sputtered to sequentially remove thin layers of material. The ejected particles are detected by a mass analyzer, which produces a mass spectrum of elements as a function of time. This raw data is then converted to give the profile of dopant [29].

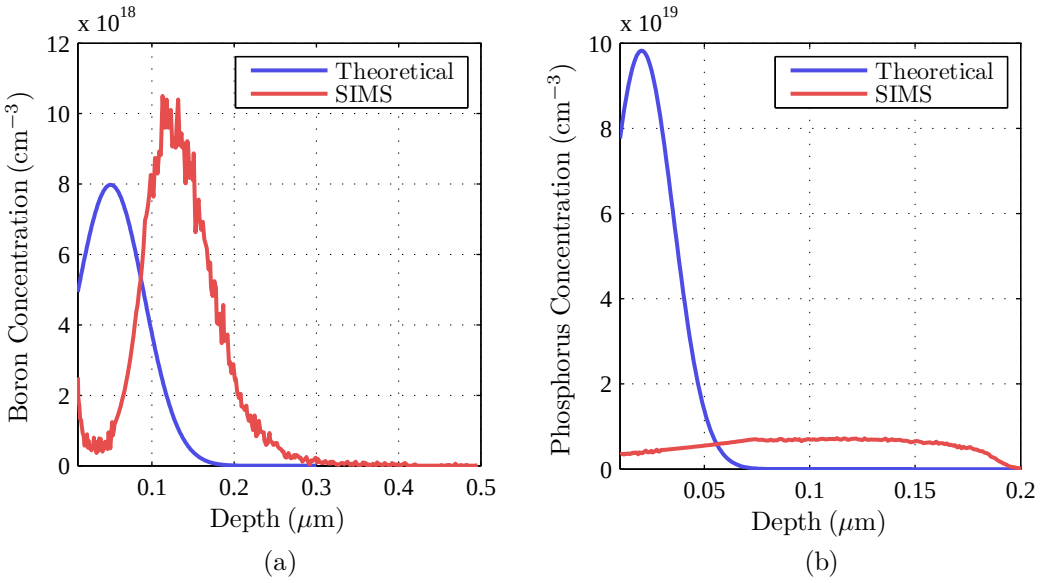


Figure 6.12: The profile of boron and phosphorus for the first batch of wafers calculated using equation *which does not include the effects of dopant diffusion during annealing* 6.1. SIMS measurements were conducted *after annealing* at 600 °C for 10 seconds. The effects of rapid phosphorus diffusion are clearly seen.

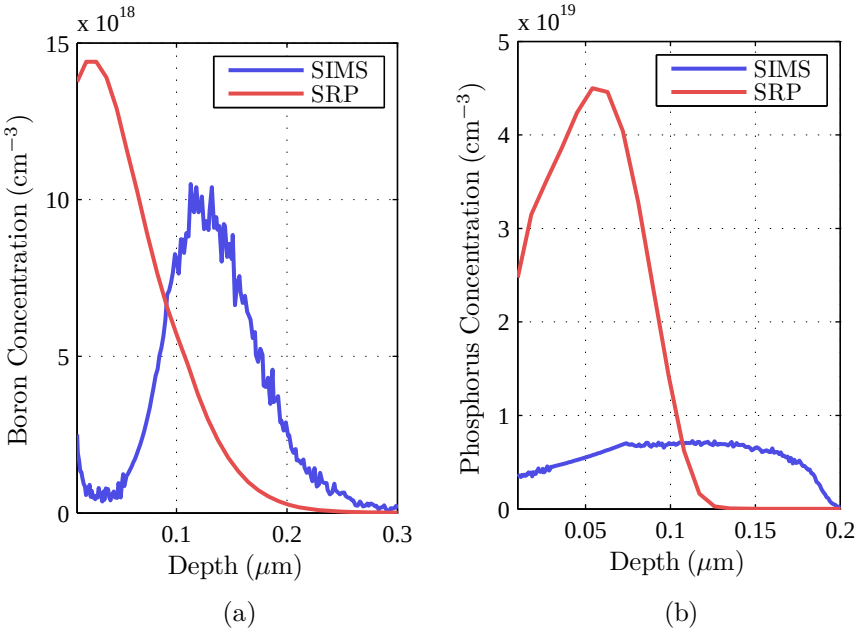


Figure 6.13: The electrically active dopant profiles for boron and phosphorus for the first batch of wafers measured using SRP. Both profiles were measured after the wafers were annealed at 600 °C for 10 seconds.

the ones predicted using SIMS, and (2) the profiles do not look remotely the same! The contradiction in the first case is well documented [8, 9, 33], and is attributed to the inaccuracy in raw data conversion due to the lack of well-calibrated standards in germanium. However, we could not explain why the profiles have completely different curvatures. Various publications on dopant implantation in germanium did not suffer from this problem [8, 9, 28, 33]. It is highly likely that the SRP profiles are wrong, since SIMS can accurately measure the concentration of dopants in germanium [33]. Errors in SRP profiles could be due to the incorrect assumption that the dopant profile measured on a bevelled surface is identical to a vertical dopant profile, or because the bevel angle is was not machined to specification, which leads to erroneous depth profile conversion. In any case, the concentration of electrically active dopants in the first batch of wafers would be less than the profiles measured using SIMS. This is also the reason why the implant doses were increased when implanting the second batch of wafers.

The doping profiles for the second batch of wafers was not analyzed using SIMS and SRP since high levels of dopant activation can be inferred from the results of sheet resistance measurements (table 6.7).

6.8 Surface Doping: Epitaxial Regrowth

To reduce the dependency on sophisticated fabrication equipments to create highly doped surfaces on the modulator (e. g., the ion implanter), two alternative fabrication methods were explored: the solid and liquid phase epitaxial regrowth [6, 27]. The latter is commonly known as alloying, a technique primarily employed in the early fifties and sixties to produce pnp junction transistors [2]. Both solid phase epitaxy (SPE) and alloying require only two steps to form highly doped layers: a metal deposition step followed by a subsequent annealing procedure in chambers evacuated to pressures below 10^{-5} mBar. The use of either method enables several processes to be streamlined, where low resistance contacts and highly doped junctions can be produced simultaneously; as opposed to the four step procedure needed in standard silicon wafer fabrication (i. e., ion implantation, RTA, electrode deposition and low temperature sintering [40]). Unfortunately, SPE and alloying have only been succesful at forming highly doped p-type layers in germanium. Data on n-type dopant incorporation is non-existent for the alloying method, while only a single reference is available for SPE [24]. Nevertheless,

we did manage to produce a functional modulator using SPE.

The apparatus used in alloying and SPE experiments, along with results from work carried out are presented in the following sections.

6.8.1 Purpose-Built Thermal Evaporator and Annealer

A rig that has a separate annealing and evaporation chamber was built in-house to conduct alloying and SPE experiments. The annealing rig can be heated to temperatures up to 750 °C (the highest temperature we have ever tried), and evacuated to pressures down to 10^{-6} mBar. The additional evaporation rig, on top of existing evarporators in the cleanroom, was necessary for evaporating toxic metals such as antimony and, in the future, arsenic onto the surface of germanium for n-type doping experiments. The construct of the rig is shown in figure 6.14.

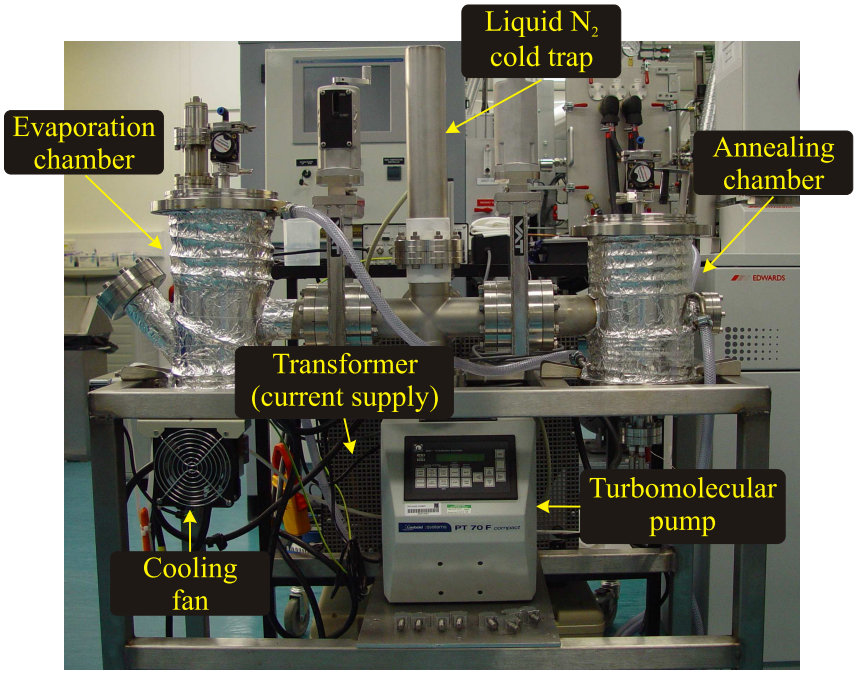


Figure 6.14: In-house built mini thermal evaporator and annealer.

Both chambers are evacuated using an integrated pumping unit that contains a diaphragm backing pump and a turbomolecular pump. The pumping system is connected to each chamber via a 4 way knife-edge conflat (CF) cross. Two valves are fitted on the CF cross to enable each chamber to be evacuated or brought up to air separately. Although a good vacuum can be obtain after several hours of evacuation

($2 - 4 \times 10^{-6}$ mBar), it could not be maintained when either rig is in operation. A liquid nitrogen cold trap was subsequently mounted to the CF cross, which significantly improved the vacuum in both chambers ($1 - 2 \times 10^{-6}$ mBar during operation). The pump-down time to reach 3×10^{-6} mBar was also reduced from several hours to just under 30 minutes. A transformer with switchable secondary voltage between 5 and 30 V, capable of supplying currents from 6.66 to 200 A is used to power both chambers. The annealing chamber is heated using a coiled graphite heating element; samples (10 by 10 mm^2) are held in graphite mounts placed 30 mm above the heater.

6.8.2 Liquid Phase Epitaxy (Alloying)

Steps involved in producing a doped junction using the alloying method are depicted in figure 6.15. An indium button is placed on an n-type germanium wafer, which is then heated to initiate the alloying process (While indium is used to form p-type surfaces in this example, other metals can be used. See table 6.3). When the indium melts, it dissolves an amount of germanium determined by the volume and temperature of the indium. Then, as the structure cools, germanium first grows from the molten drop onto the wafer. This layer of germanium is very thin, and contains a considerable concentration of indium ($\approx 4 \times 10^{18}$ cm^{-3}). The remaining part of the solidified melt consists mainly of indium. If the supply of doping material is limited, then the depth of alloying is dependent of the temperature and the wetted area (by molten indium). If the volume of charge is equal or exceeds the volume of germanium, then the alloying time must be controlled to prevent the entire sample from fusing completely [13, 40].

The main parameters that influences the density of dopants in regrown regions are the peak alloying temperature and the cooling rate. The former is inferred from data in table 6.3, where the solid-solubility limit for different acceptors in germanium only peak at higher temperatures. To prevent the out-diffusion of dopants during recrystallization (since the solid-solubility limit reduces with temperature), the sample must be cooled down rapidly at rates of $5 - 20$ °C/min [22, 27]. Slower cooldown rates (< 3 °C/min) could reduce the peak doping density by up to one order of magnitude [27].

Junctions produced by the alloying method are usually abrupt, provided that the dwell time at the alloying temperature is kept short to prevent a large number of ions from diffusing into the wafer. The uniformity of junctions formed depends on the crystallographic orientation of the wafer, since the rate of penetration is different for the various crystal planes. (1,1,1) oriented wafers produces the most homogeneous

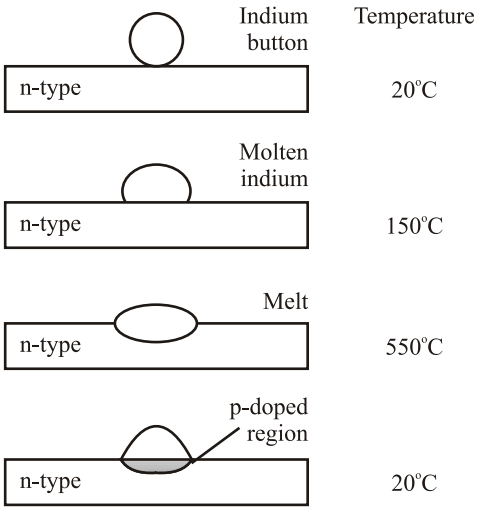


Figure 6.15: Steps in the production of an alloyed junction

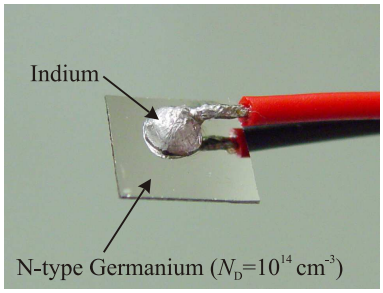


Figure 6.16: Indium bead alloyed to a $10 \times 10\,\mathrm{mm}^2$ sample that is $300\,\mu$m thick.

penetration depth because this plane is most resistant to alloying due to way atoms are oriented in this plane [22].

Temperature (°C)	Indium $\times 10^{18}$ (cm^{-3})	Gallium $\times 10^{20}$ (cm^{-3})	Aluminium $\times 10^{20}$ (cm^{-3})
400	1.5	3.5	—
500	2.1	4.5	4.0
600	2.9	4.9	4.1
700	3.5	5.0	4.1
800	4.0	4.1	3.9

Table 6.3: The solid solubility limit of different acceptors in Germanium [37]

Junctions with higher electrically active dopant concentrations can be obtained by mixing indium with gallium ($N_D \approx 10^{19}\,cm^{-3}$), or by using aluminium in place of indium ($N_D > 10^{20}\,cm^{-3}$) [23, 27]. We have performed alloying experiments using In and Al beads ($1-2\,mm$ in diameter) on thin slices of n-type germanium to form pn diodes (figure 6.16). This method enables the determination of p-type juction formation using simple current and voltage measurements. The surfaces of the samples are cleaned and etched as described in section 6.6, and metal beads are dipped in HCl to remove any oxide layers present. Table 6.4 lists all the alloying parameters used in experiments along with results from pn structure verification measurements.

For samples with $10^{17}\,cm^{-3}$ average donor concentration, only aluminium alloyed at 600 °C was successful at forming diodes (table 6.4). Highly doped wafers were used to gauge the density of acceptors incorporated into the wafer. If diodes are formed, it could be caused by the compensation of donor ions by acceptors ($N_A > 10^{17}\,cm^{-3}$), or because donor ions are somehow ejected from the sample, to be replaced by aluminium ions. The former is more likely, since wafer compensation is routinely use to change the resistivity of germanium [36]. This shows that indium is ineffective at introducing high peak acceptor density in germanium, since it was not able to convert the regrown region of the n-type wafer with a donor density of $10^{17}\,cm^{-3}$ to p-type.

Note also that a diode was formed using aluminium heated to 310 °C. This is below the melting point of aluminium thus alloying was not responsible for the formation of pn junction. Also, the diffusion of aluminium into germanium can be ruled out since the diffusion coefficient is negligible at that temperature ($<10^{-21}\,cm^2/s$ [23]). However, the

Metal	Alloying Temperature (°C)	Cooling rate (°C/min)	Wafer Doping (cm^{-3})	Diode
In	350	10	10^{14}	Yes
	350	20	10^{14}	Yes
	350	10	10^{17}	No
	350	20	10^{17}	No
	450	10	10^{17}	No
	450	20	10^{17}	No
	550	10	10^{17}	No
	550	20	10^{17}	No
	650	10	10^{17}	No
	650	20	10^{17}	No
In/Al/In*	550	10	10^{17}	No
	550	20	10^{17}	No
Al	310	10	10^{14}	Yes
	310	10	10^{17}	No
	450	20	10^{17}	No
	600	10	10^{17}	Yes
	600	20	10^{17}	Yes

Table 6.4: Different alloying parameters and metals used for doping n-type germanium wafer to form pn diode. Samples were all alloyed for 10 minutes, sufficient for the melt to be saturated with germanium. (*Al film sandwiched between two In films).

transport in the opposite direction is much more rapid, since the diffusion coefficient of germanium in aluminium is $\approx 10^{-11}\,\mathrm{cm}^2/\mathrm{s}$. The diode was theferfore created by the precipitation of germanium ions from aluminum, which forms a highly doped p-type layer on the substrate. The solid phase dissolution of germanium ions in the metal layer is the basis for the growth of doped layers using SPE. This is covered in the next section.

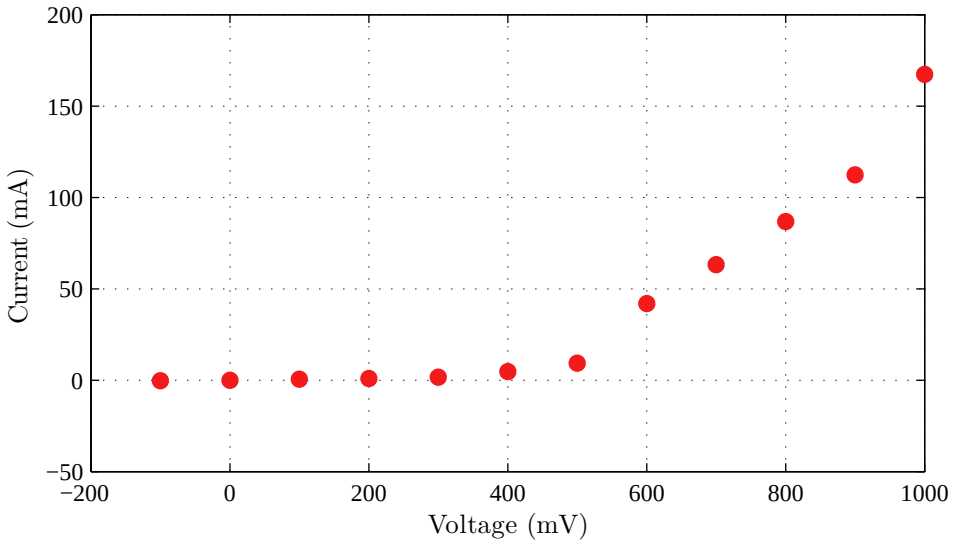


Figure 6.17: An example of a pn diode produced using aluminium beads alloyed onto n-type germanium wafer with donor concentration of 10^{17} (cm^{-3}). The structure is similar to the one shown in figure 6.16

To utilize the alloying technique for producing the modulator, the source of dopants must be deposited using lithography so that it may be accurately placed above the electrode (remember that it is regions under the electrode that needs to be highly doped). In fact, the source itself can double as the electrode since it is made of metal. Therefore, only thin metal layers deposited using an evaporators is used for alloying.

We were unsuccessfull at alloying thin indium pads to germanium because of its poor wetting [22]. Once the metal reaches melting point, surface tension causes the bulk of material to form a blob (figure 6.18). Although better wetting properties can be achieved by using solder flux [22], its use is undesirable since cross contamination and damage to the modulator may result.

Due to time constraints, all work on alloying was stopped to focus on using solid-phase epitaxial regrowth (SPE) to produce the modulator. SPE has several significant

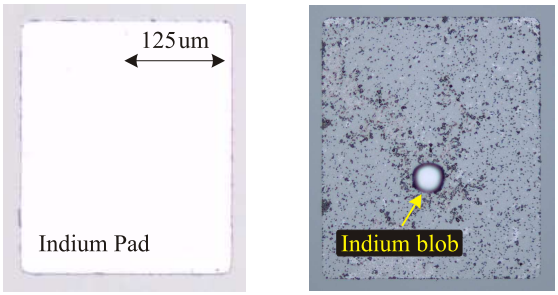


Figure 6.18: (a) A 1 μm thick In pad deposited on the surface of germanium, and (b) an indium blob formed due to surface tension when the sample was heat treated at 300 °C.

advantages over alloying. It is a low thermal budget technique that operates below the melting temperature of the dopant material. This reduces the problem of unwanted impurity diffusion into the modulator at elevated temperatures. Dopant incorporation via SPE is also independent of the cooling rate, unlike the alloying method which requires the sample to be cooled quickly to retain the doping level in the regrown region. As samples are allowed to gradually cool to room temperature (using SPE), the regrown region would be more robust and contains less defects. Finally, there is evidence from the literature to shown that injecting n-type layers have been formed using SPE [24]. (Although there is no literature available to support the use of alloying to fabricate n-type layers, we had in mind to alloy tin doped with arsenic or antimony with germanium to form such layers (tin by itself is a neutral impurity). This method may not be effective since the SSL of tin in germanium is several times greater than that of antimony and arsenic [37].)

6.8.3 Solid Phase Epitaxy

The general steps to introduce dopants into germanium using SPE are very similar to alloying [6]. A metal layer (aluminium for acceptors and antimony for donors) is first deposited onto the surface of a clean germanium sample. The sample is then heated to temperatures below the eutectic temperature for several minutes to allow the germanium to dissolve into the metal film (Ge-Al eutectic, 424 °C; Ge-Sb eutectic, 590 °C [24]). Upon cooling, the dissolved germanium will precipitate out of the metal to epitaxially grow on the surface of the sample. For this reason, it is vital that interfacial layers (e. g., native oxide and dirt) are removed prior to metal deposition

or it will prevent the semiconductor from dissolving into the metal layer in the first place. The thickness of the regrown layer depends on the solid solubility limit (SSL) of germanium in the metal at the annealing temperature. Table 6.5 lists the SSL for germanium in Al. No SSL data could be found for germanium in antimony (though the reverse is available [32]).

Temperature (°C)	Solid Solubility Limit (cm^{-3})
100	3.09×10^{17}
200	1.05×10^{20}
300	2.11×10^{20}

Table 6.5: The solid solubility rate of germanium in aluminium

The parameters used in SPE experiments are listed in table 6.6. PN junctions were successfully produced using each configuration. To the best of our knowledge, we are the first to produce n-type injecting contacts on germanium using a single layer of Sb. Similar experiments were conducted by Ottaviani back in the early seventies without success [24]. Failure in the past could be a result of poor vacuum conditions incurred by the use of an oil diffusion pump. These pumps are notorious for contaminating vacuum chambers with hydrocarbon vapours due to the back-streaming of oil vapour [17].

After several successful attempts at producing n-type junctions, efforts were immediately diverted to the fabrication of a pin diode. This was done to establish the basis of using SPE in producing both n and p-type junctions.

A pin diode was succesfully fabricated via the following procedure: (a) deposition of the front and back side of a 1 mm thick, 10 by 10 mm^2 intrinsic germanium sample with Al and Sb film respectively (both 1 μm thick), (b) anneal the sample at 280 °C for 15 minutes, and (c) cool the sample to room temperature at a rate of 6 °C /min. The electrical performance of the diode measured using simple probes is shown in figure 6.19.

The thickness of the junctions formed can be estimated with the help of table 6.5. At 280 °C, the solid solubility limit of Ge in Al is roughly 2×10^{20} cm^{-3}. This means that the total germanium dissolved in the 1 μm thick Al film is roughly 8 nm thick. If 90% of the dissolved germanium precipitates out from the aluminium upon cooling [6], then a p-layer with a thickness of ≈ 7 nm is expected. Similarly, the regrown layer using Sb is expected to be very thin.

Metal	Dwell Temperature (°C)	Dwell Time (mins)	Wafer Doping (cm^{-3})
Al (acceptor)	350	15	10^{14}
	300	30	10^{14}
	300	15	10^{14}
	300	10	10^{14}
Sb (donor)	280	60	10^{15}
	280	30	10^{15}

Table 6.6: Different SPE parameters and metals used for doping germanium samples to form a pn diode. N-type wafers were used when acceptors were incorporated while p-type wafers were used when donors were added. All of the configurations above produced pn diodes. The thickness of the Al and Sb films deposited are 1 and 0.2 μm respectively. A temperature ramp of 6 °C /min was used to raise the temperature to the set point. The same rate was also used to cool the samples.

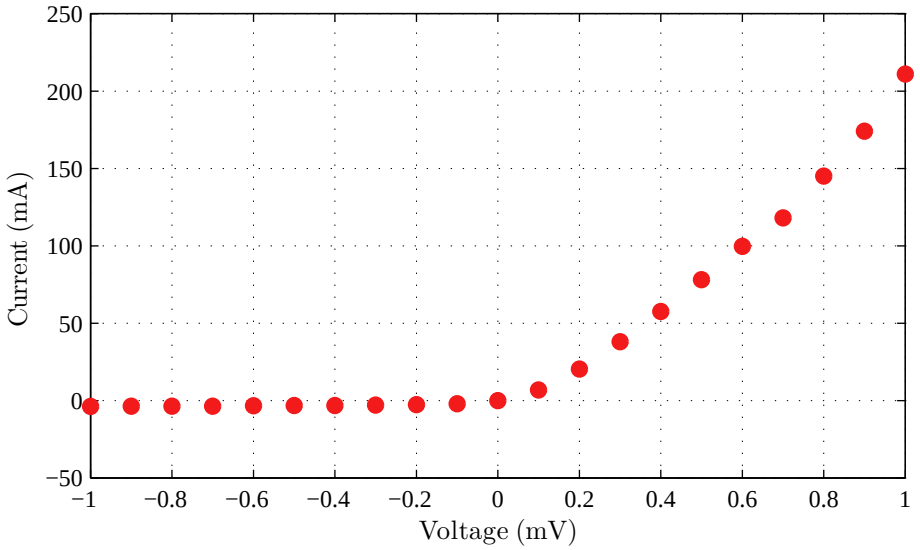


Figure 6.19: The electrical performance of the first pin diode produced using solid phase epitaxial regrowth.

Thicker doped layers can be obtained using the multi-layer film structure shown in figure 6.20 [23]. Upon heating, the layer of amorphous germanium will dissolve into the Al film. Unlike crystalline germanium, the higher free energy of its amorphous counterpart then causes the dissolved Ge atoms to diffuse through the entire depth of the Al layer. If no interfacial layers are present, epitaxial growth can be achieved when crystalline precepitates are formed at the other end of the metal film.

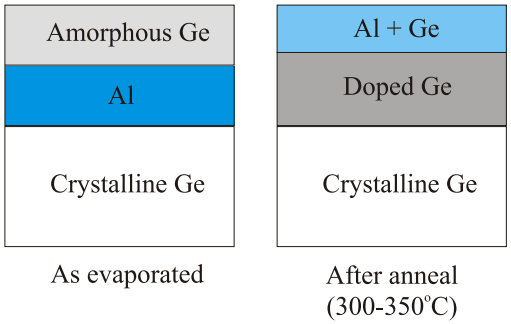


Figure 6.20: Multilayer configuration to deposit thick p-type layer on germanium using SPE.

Although this technique was not used for producing the modulator (due to time constraints), a single sample was produced to enable the measurement of acceptor concentration in the doped layer. In this sample, Al and Ge films with similar thicknesses ($\approx 65\,\text{nm}$) were sequentially deposited on the substrate at rates of 0.5 and 0.1 nm/s respectively without breaking vacuum. The sample was then annealed at 300 °C for 15 minutes and subsequently cooled to room temperature at a rate of 6 °C/s. The transmission line method (section 6.10) was then used to estimate the resistivity of the doped layer. The modified structure required for measurements to be carried out is shown in figure 6.21. Once the resistivity is known, the total acceptor concentration was deduced from resistivity versus impurity concentration data from Cuttriss [11].

A minimum acceptor concentration of $10 \times 10^{21}\,\text{cm}^{-3}$ was measured using this technique. This is in agreement with values given by Ottaviani *et. al.*, obtained using Hall effect measurements. Although the accuracy of our results may be compounded by errors in measuring the actual thickness of the doped layer ($\pm 10\,\text{nm}$ drift in thickness measurement due to vibrations in the cleanroom), the resulting acceptor concentrations is still within the order of $10 \times 10^{21}\,\text{cm}^{-3}$. Therefore, SPE is a viable alternative to dope germanium with acceptors. In fact, the achievable peak dopant concentra-

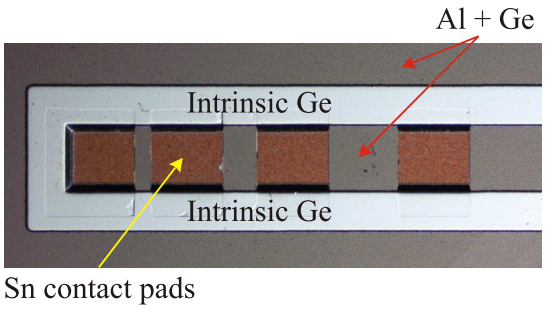


Figure 6.21: The structure to measure the concentration of acceptors in the aluminium doped germanium layer (Al+Ge) using the transmission line method (section 6.10). The doped layer is 60 nm thick, and was grown using the configuration as shown in figure 6.20 (the top aluminium layer was removed by etching in HCl to allow the Sn pads to be deposited).

tion is at least an order of magnitude higher than what is attainable through boron implantation [33].

Unfortunately, we did not succeed in creating highly doped n-type layers using Sb with the multi-layer structure. Visual inspection of the sample annealed at 280 °C for up to an hour revealed that the germanium layer did not diffuse through the Sb layer. However, Ottaviani *et. el.* succeeded in producing n-type injecting contacts by adding an extra layer of Sb to the stack (i. e., Sb/Ge/Sb/substrate), [24]. They attributed the success in forming injecting contacts to the use of the amorphous layer, since backscattering experiments showed that *limited* solid-solid reaction between the Sb and amorphous germanium layer did take place. The evidence we have (i. e., the successful fabrication of n-type layer on intrinsic germanium as listed in table 6.6) suggest that the n-doped layer was most likely formed by germanium dissolved from the substrate and not from the amorphous layer.

6.9 Dielectric Layer

Dielectric layers are needed to allow large contact pads to be deposited on the modulator without short circuiting the electrodes (figure 6.34). Modulators were succesfully produced using dielectric layers made from negative photoresist and SiO_2. Better performing modulators were fabricated using the latter because its more robust and contains fewer defects. For completeness, the deposition process for both materials are briefly described in the following sections.

6.9.1 Negative Photoresist

SU-8:2 negative photoresist (NPR) from MicroChem were used to fabricate early prototypes of the modulator. It is used because when set, it becomes a very durable non-conducting epoxy that adheres very well to the semiconductor surface and could even withstand mild acids and solvents.

The following steps were used to deposit and cure the resist: (a) depositing the NPR on the wafer using a syringe with 4.5 μm filters attached; the resist must be dripped on the middle of the sample to prevent streaking during spin coating (the opposite is true for positive photoresist), (b) place sample on a vacuum spindle and spin coat at 500 rmp at a ramp rate of 100 rpm/s; without stopping, the sample is then ramped to 3000 rpm at 300 rpm/s and left to dwell for 30 seconds. (c) the sample is then baked on a hotplate at 65 °C for one minute followed by a 95 °C bake for 3 minutes to evaporate solvents from the resist; (d) position sample in the mask aligner with the mask in place and irradiate the sample for 15 seconds at 7.2 mJ/cm^2; (e) post bake the sample for 2 minutes at 95 °C to initiate the cross linking process (this turns the photoresist into a durable epoxy) and (e) immerse sample in developer for 1 minute and rinse with isopropanol.

The resulting patterns transferred to the photoresist are shown in figure 6.22. Note that negative and positive photoresist (PPR) were used simultaneously to transfer the pattern of the electrode to the modulator. The PPR is stripped away after a subsequent metal layer is deposited, leaving a large pad of metal on top of the NPR and only a narrow stripe of metal in direct contact with the sample (i. e., the electrodes). Although modulators can be produced using this method, contacts formed were highly resistive. At 2 V forward voltage, a modulator with structure similar to one listed in table 5.1 only managed to conduct 0.1 A of current. This could be caused by thin resist residues left on the surface of the sample, which form a resistive layer between the electrode and the substrate. A large number of pin holes and defects were also present on the resist layer. For these reasons, NPR was later replaced with SiO_2, since it does not suffer from the same problems.

6.9.2 Silicon Dioxide

Silicon dioxide layers were deposited on the modulator using plasma enhanced chemical vapour deposition (PECVD). It was done commercially at Phillips Reseach Eindhoven. Because only thin layers of metal can be deposited on the modulator to form the

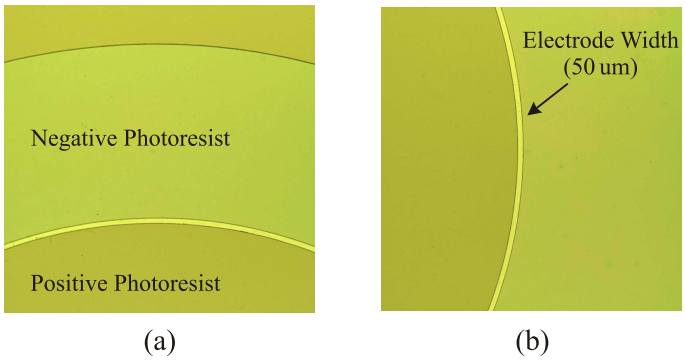


Figure 6.22: Patterns transferred to the negative and positive photoresist

electrodes (<2 μm), the thickness of the oxide layer was limited to 300 nm. If the thickness of the dielectric layer exceeds that of the metal layer, the electrode will not be in electrical contact with the contact pads located on the dielectric layer (figure 6.34).

To form the electrodes on the modulator, windows must be etched in the oxide layer (figure 6.34). This is done by defining patterns on a layer of photoresist that covers the dielectric layer, followed by immersion in HF acid to etch away regions of the oxide not protected by the photoresist. Diluted HF was used to prevent the surface of the substrate from being corroded once the oxide layer was etched away (figure 6.9). This procedure was problematic because the layer of photoresist constantly peeled off after soaking in 10:1 H_2O:HF for 10 minutes (15 minutes was required to etch away 300 nm of SiO_2). The problem was initially thought to be caused by bad adhesion, thus measures to enhance the adhesion of photoresist like the use of adhesion promoters (TI Primer) and by hard baking the photoresist were performed [40]. The peeling of photoresist persisted. A collegue later pointed out that undercutting of the SiO_2 layer was responsible for the peeling of photoresist, since this problem is regularly encountered in standard silicon wafer processing. The problem was solved using 1:1 H_2O:HF to etch the oxide layer. To prevent the surface from being damaged, the sample must be retrieved immediately after the patterns are etched. In our samples, only 15 seconds was required. The pattern of the electrode that was transferred to the dielectric layer is shown in Figure 6.23

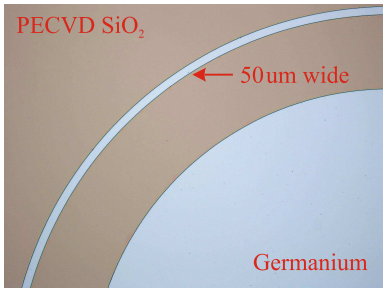


Figure 6.23: The pattern of annular electrode and the aperture transferred to the dielectric layer.

6.10 Ohmic Contacts

This section details the last procedure involved in the production of the *core* modulator structure: ohmic contact formation. Low resistance contacts are crucial to ensure that high current modulation can be achieved without excessive power dissipation. From simulations performed, the default structure is predicted to dissipate less than 2 W of power to achieve $\approx 95\%$ depth of modulation at 10.59 μm (figure 5.8). Ohmic contacts were assumed in simulations. In real devices however, the actual power dissipation is often higher due to the presence of parasitic resistance (e. g., resistive interfacial layer) and large potential barrier caused by the difference in work function between the metal and the substrate (section 3.4.1). Efforts to reduce the latter is often compounded by surface states at the interface that prevent the use of different metals to engineer low barrier heights. Nevertheless, by doping the surfaces below the electrodes, the tunnelling probability of carriers can be increased to reduce the effective contact resistance. The deleterious effects of interfacial layers can be mitigated through careful surface preparation (section 6.6), though a thin layer of oxide will always be present unless in-situ wafer cleaning used. A solution to remove the oxide layer *after* the electrode has been deposited is to dissolve it in the metal layer by heating the sample. Both alloying and SPE are viable methods to achieve this.

Low resistance contacts were succesfully produced using the surface preparation procedure detailed in section 6.6. Even better contacts were obtained by heat treating the samples after annealing. The properties of contacts formed were analyzed using the so called 'transmission line method' (TLM). Therefore, the basics of TLM is first reviewed.

6.10.1 The Transmission Line Method

The structure depicted in figure 6.24 is used in TLM measurements. Three metal pads with an area defined by w × d, where w and d are respectively the width and length of the pads, are deposited on a *doped* semiconductor surface. The separation between the pads is set by l_1 and l_2.

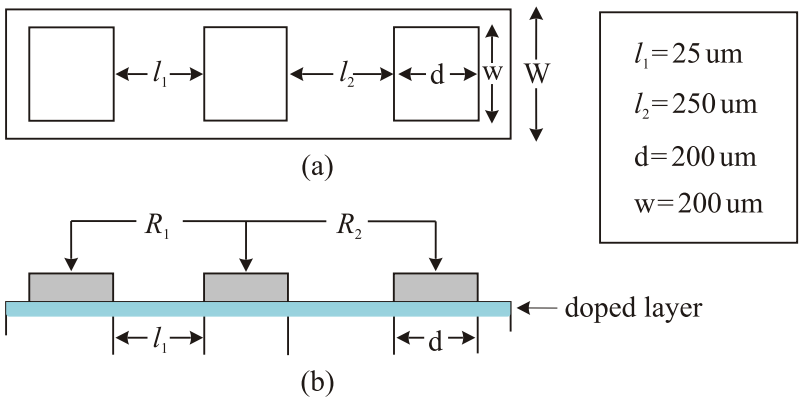


Figure 6.24: The structure for determining the total contact resistante R_c using TLM.

By measuring the resistance between adjacent contacts using needle probes (R_1 and R_2), the resistance of each contact R_c (Ω) can be evaluated using:

$$R_c = \frac{R_2 l_1 - R_1 l_2}{2(l_1 - l2)} \tag{6.2}$$

Note that R_c is assumed to be the same for each contact and that the current is confined within the doped layer when R_1 and R_2 are measured. Once R_c is known, the specific contact resistivity ρ_c ($\Omega \cdot \mathrm{cm}^2$) can calculated via the equation

$$\rho_c = R_c \cdot A, \tag{6.3}$$

where A is the area of the contact pad. ρ_c enables the quality of contacts with various shapes and sizes to be compared since it is independent of contact area.

For the properties of the contact to be analyzed accurately using TLM, several criteria must be satisfied. Firstly, the flow of current must be confined within the doped region. In addition, the doped layer must be very thin or the effective contact resistance will be higher because carriers have to travel further before exiting the electrode (figure 6.25b) [3]. The samples we used had very thin doped layers (<200 nm) created using ion implantation; thus the second requirement is satisfied. To confine the current within the doped layer, the standard method (and the only one the author

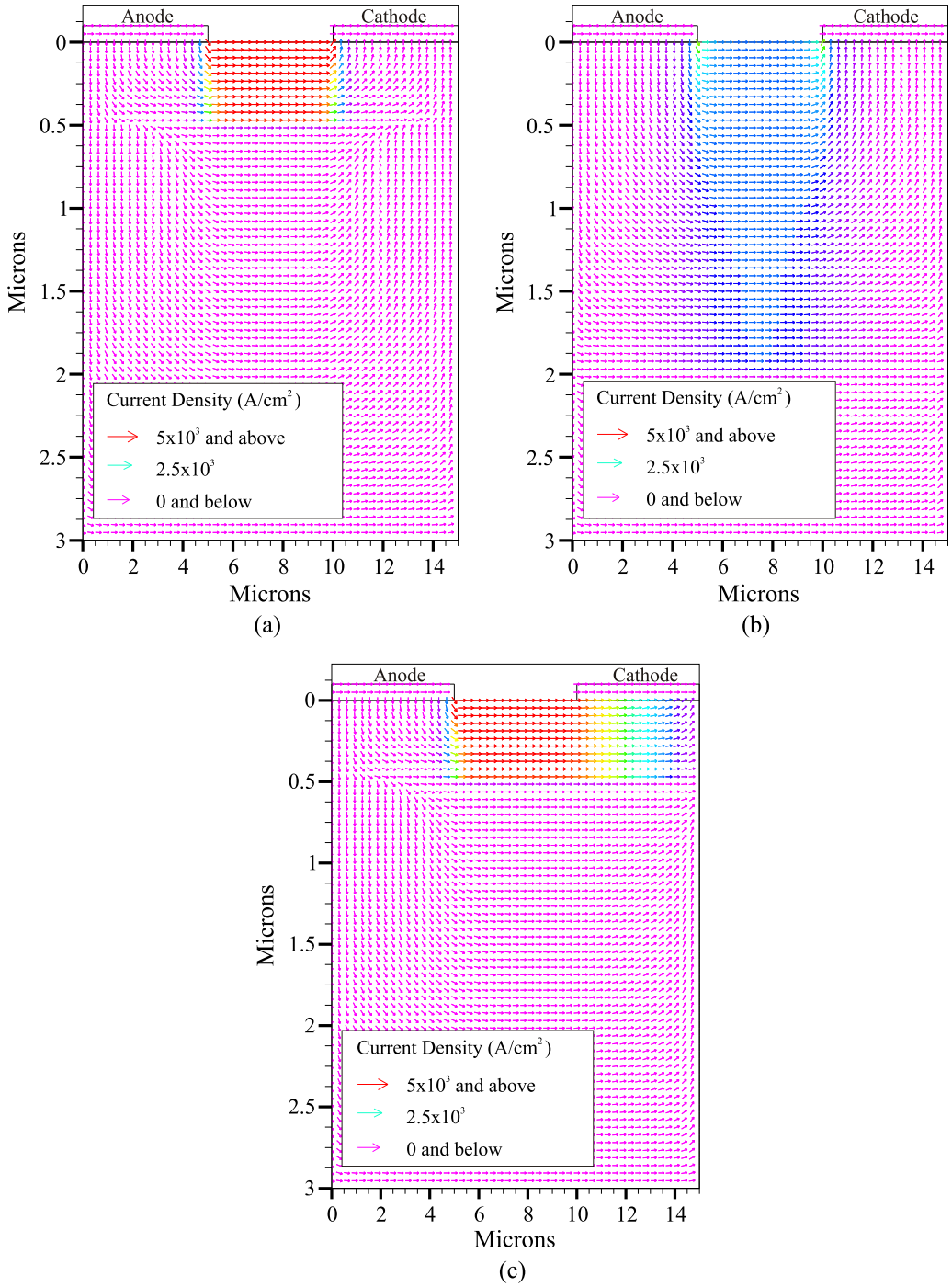


Figure 6.25: A current flow of 2.6 A between two electrodes simulated using ATLAS. The current is confined within the doped region, which assumes a stepped profile and has a density of $10^{19}\,\mathrm{cm}^{-3}$. The respective dopant depth and ρ_c for the three sub figures are (a) $0.5\,\mu$m and $10^{-6}\,\Omega\cdot\mathrm{cm}^2$; (b) $2\,\mu$m and $10^{-6}\,\Omega\cdot\mathrm{cm}^2$; (c) $0.5\,\mu$m and $10^{-2}\,\Omega\cdot\mathrm{cm}^2$. For contacts with high resistance, current crowding will force the current to flow further along the doped region before exiting the cathode.

found) is to use pn junction structures [29]. For a p-type layer, the substrate should be n-type and vice versa. However, for TLM results to be transferable to the production of the modulator, intrinsic samples with doped surfaces must be used, since the modulator is essentially a pin diode with electrodes attached to the 'p' and 'n' doped surfaces. Since no data are available on the use of intrinsic samples in TLM measurements, the flow of current in the aformentioned structure was analyzed using ATLAS. The results are shown in figure 6.25. As can be seen, the movement of carriers are confined within the doped layer without using the pn junction configuration. This is not surprising since the resistivity of the doped layer ($<0.05\,\Omega\,\mathrm{cm}$) is several orders of magnitude lower than the resistivity of intrinsic germanium ($>50\,\Omega\,\mathrm{cm}$).

The second criteria to be satisfied is that the length of metal pad 'd' should be sufficiently long to mitigate the effects of current crowding. If the length is too short, current crowding will result since carriers are forced through the small contact area. If the pad was larger, current can exit through the path of least resistance without crowding (figure 6.25c). Mathematically, the behaviour of contact resistance can be expressed as [3]:

$$R_\mathrm{c} = Z \coth(\alpha d), \tag{6.4}$$

where Z is the characteristic line resistance, and α the attenuation constant. Z and α is given as

$$Z = \frac{1}{w}\sqrt{R_\mathrm{sh} \cdot \rho_\mathrm{c}} \tag{6.5}$$

and

$$\alpha = \sqrt{\frac{R_\mathrm{sh}}{\rho_\mathrm{c}}}. \tag{6.6}$$

R_sh is the sheet resistance of the doped layer ($\Omega/\square$). As αd decreases beyond 1.5, R_c rises sharply. But if the length is sufficiently long to increase the product of αd to ≥ 3, then the value of R_c will approach Z asymtotically. The minimum contact length required is estimated using equation 6.6. Since a contact is considered ohmic if its ρ_c is less than 10^{-6} [19], and because the sheet resistance for shallow implants in germanium is usually less than $250\,\Omega/\square$ [9], a pad with minimum length of $2\,\mu$m is needed. We have used $200 \times 200\,\mu\mathrm{m}^2$ pads TLM measurements, which is large enough for contact resistivity as high as $1.1 \times 10^{-2}\,\Omega \cdot \mathrm{cm}^2$.

The third criteria for accurate ρ_c measurement is to ensure that the contact width 'w' is wide enough to cover the entire width of the conducting stripe 'W' (i. e., the doped layer; see figure 6.24). If w<W, lateral current crowding will result (figure 6.26).

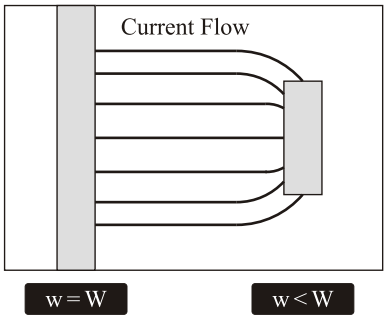


Figure 6.26: Lateral current crowding at a contact with dissimilar width (w) [3].

In figure 6.24, 'w' was intentionally drawn smaller than 'W', to highlight a common hurdle in fabricating TLM structures [3]. We overcame this problem by depositing pads with w>W, followed by a mesa etch with inner width 'W' that encompasses all three pads. The resulting structure is shown in figure 6.27. As can be seen, the width of the metal pads is similar to the conducting stripe. It is worth pointing out that this step also allowed us to form the conducting stripe simulateously, since the substrate we used had the entire surface implanted with dopants.

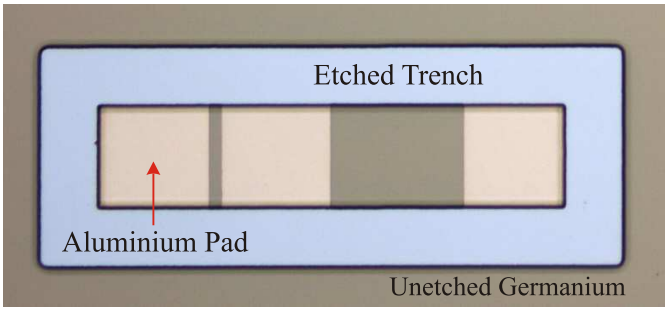


Figure 6.27: Fabricated resistor test structure for determination of contact resistance (R_c).

Finally, R_1 and R_2 must be measured within the range where the current and voltage behaves linearly (see figure 6.28). Non-linear IV behaviour is attributed to current crowding which invalidates the value of resistance measured.

Apart from ρ_c, the sheet resistance of the conducting stripe can be measured. The total resistance between two pads can be expressed as:

$$R_{\mathrm{T}} = 2R_{\mathrm{c}} + R_{\mathrm{sh}}\left(\frac{l}{w}\right), \tag{6.7}$$

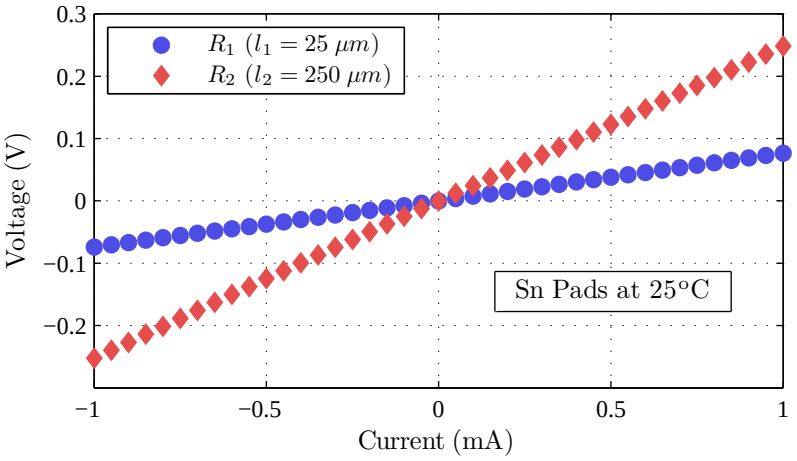


Figure 6.28: Current and voltage curves measured for $R1$ and $R2$ using as deposited Sn pads.

which is the resistance of two contact pads plus the resistance of the doped layer. If two of such measurements were made on different pairs of pads (i. e., R_1 and R_2), then the sheet resistance can be equated to:

$$R_{\text{sh}} = (R_1 - R_2) \cdot \frac{w}{l_1 - l_2}. \tag{6.8}$$

If R_{sh} is known, the density of dopants in the doped layer can be determined by multiplying R_{sh} with the thickness of the doped layer t to obtain its resistivity ρ ($\Omega\,\text{cm}$). ρ can then be converted to the effective doping density using the published data of resistivity versus doping concentration [35]. The doping profile must be be stepped and not graded else the doping concentration obtain is just an average through t. The density of acceptors in the doped layer fabricated using SPE were determined using this method (figure 6.21).

6.10.2 Sample Preparation

Intrinsic germanium samples with surfaces implanted with dopants were used as substrates to fabricate TLM structures. The samples were sawn from the first and second batch of ion implanted wafers (see table 6.2) into $10 \times 10 \times 1\,\text{mm}^3$ pieces. Thus there is a mixture of doping profiles in the samples used. Prior to metal deposition, samples were cleaned using the procedure described in section 6.6.

6.10.3 Experimental Results

The properties of contacts formed on germanium for different metals and annealing temperatures are shown in table 6.7. (Other configurations were produced but only results that are reliable and consistent are listed). In and Al contacts were deposited on p-type surfaces while Sn electrodes were formed on n-type surfaces.

Electrode	Annealing Temperature (°C)	R_{sh} ($\Omega/\square$)	R_c (Ω)	ρ_c ($\Omega \cdot cm^2$)	R_{mod} (Ω)
In	120	197	85.96	3.43×10^{-6}	6.74×10^{-3}
In	200	182	36.09	1.44×10^{-6}	2.83×10^{-3}
Sn	none	152	42.6	1.7×10^{-6}	3.34×10^{-3}
Sn	200	119	43.86	1.75×10^{-6}	3.44×10^{-3}
Sn (SPE)**	300	166	35.89	1.43×10^{-6}	2.81×10^{-3}
Al	none	177	47.07	1.88×10^{-6}	3.69×10^{-3}
Al	350	198	23.76	9.50×10^{-7}	1.86×10^{-3}

Table 6.7: The contact properties for different metals and annealing temperature. **The conductive stripe is doped with aluminium using SPE and is 60 nm thick.

All of the samples used have sheet resistances of less than $200\,\Omega/\square$[7]. If the depth of the implanted layer is ≈ 150 nm (a very conservative estimate for acceptors but a reasonable one for donors; see figure 6.13), the doping level for both species of dopants are in the range of 3×10^{18} to $6 \times 10^{18}\,cm^{-3}$. This however is just the average concentration because the implanted profile is graded and not stepped. Thus the peak doping concentration should be higher—especially for acceptors. One could safely conclude that the peak dopant concentration in the second batch of wafers not measured using SIMS and SRP should exceed $6 \times 10^{18}\,cm^{-3}$ ($\geq 10^{19}\,cm^{-3}$ is needed for the modulator to perform well electrically and optically).

For Al and Sn electrodes, low resistance contacts can be obtained even without performing the post annealing step to dissolve the interfacial oxide layer ($\rho_c < 10^{-6}\,\Omega \cdot cm^2$

[7] because Sb pads are very brittle, they were easily damaged during measurements thus the contact properties listed in table 6.7 for Sb is incorrect

is considered ohmic [19]). Thus it can be inferred that the procedure used for cleaning the samples is sufficient to remove most of the interfacial layers except for the ubiquitous film of germanium oxide. Annealing at temperatures between 280 °C to 350 °C further improved the quality of contacts formed. Therefore, Al and Sn contacts (with additional heat treatment at 280 °C for 15 minutes) were used as the anode and cathode for the modulator respectively.

With the fabrication of ohmic contacts reviewed, every step needed to produce the core modulator structure is now covered. Before illustrating the main steps involved in making the modulator, the usefulness of a single layer antireflection coating is briefly presented in the next section.

6.11 Anti-Reflection Coating and Surface Passivation

Due to the high refractive index of germanium ($n \approx 4$ from $3 - 14\,\mu$m), anti-reflection coatings (AR) must be deposited on the modulator to reduce its reflection losses. If its operating window spans several micrometers (8 to 14 μm), multilayer AR coatings are needed to homogeneously increase the transmittance across the region of interest [20]. Though such coatings are required by the modulator for it to be used in pyroelectric detector applications (e. g., gas sensing [15] and thermal imagers), it is beyond the scope of this project to produce such layers since they can be made commercially. However, because it has been shown that low SRV values 4.1.6 can be preserved by passivating the surface of the modulator with ZnS [14], and since ZnS can be used as a single layer AR coating for germanium, initial work to integrate the use of ZnS layers as an AR coating and as a passivation layer for the electrically driven modulator was conducted. AR coated germanium samples with wavelength centred at 10.59 μm were produced using a thermal evaporator. Before presenting the results, equations used in modelling the properties of a single layer AR coating on germanium as a function of incident electromagnetic wave (EM) is first reviewed (all equations used were taken from standard optics textbooks [4, 20]).

6.11.1 The Transmission Through AR Coated Germanium

The passage of EM wave through the a stratified medium with two interface (i. e., AR coated germanium) is shown in figure 6.29. The light travels through three medium consecutively, starting from the air, then in ZnS and finally through germanium. The

corresponding properties for each layer are labelled with subindex 1, 2, 3 respectively. For example, the refractive index and the angle of incidence of EM wave on the first layer are n_1 and θ_1. To AR coat the layer of germanium at wavelength λ_i, the reflection from its surface and from the ZnS layer must interfere destructively at that wavelength. This means that the phase difference between the aformentioned layers must be π. To achieve this, the thickness of the ZnS layer h must be equal to $a[\lambda_i/(4n_2)]$, where $a = 1, 3, 5 \cdots$. As 'a' increases, the interference fringes formed from the superposition of light reflected from layer 1 and 2 also increase, making the AR coating highly wavelength dependent. Therefore, 'a' of 1 was used in the production of AR coating to retain a broad fringe at λ_i (figure 6.30).

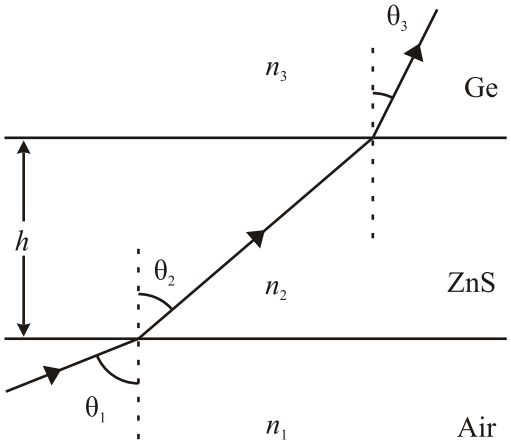


Figure 6.29: Propagation of an electromagnetic wave through a homogeneous film.

Two light polarizations are considered in estimating the transmittance through the stratified layer: transverse electric mode (TE) and transverse magnetic (TM) mode. In TE mode, the wavevector is perpendicular to the incidence plane, while in TM mode, it is polarized parrallel to the incidence plane. The transmitted intensity of light for both modes are equal at normal incidence. In TE mode, the absolute transmission (T_total) and reflection (R_total) through the stratified medium can be expressed as:

$$T_\mathrm{total} = \frac{\rho_3}{\rho_1}|t|^2 \tag{6.9}$$

and

$$R_\mathrm{total} = |r|^2, \tag{6.10}$$

where ρ is equal to $n\cos(\theta)$. 't' and 'r' are respectively the reflection and transmission

coeffecient given as:

$$r = \frac{(m_{11} + m_{12}\rho_3)\rho_1 - (m_{21} + m_{22}\rho_1)}{(m_{11} + m_{12}\rho_3)\rho_1 + (m_{21} + m_{22}\rho_1)} \tag{6.11}$$

and

$$t = \frac{2\rho_1}{(m_{11} + m_{12}\rho_3)\rho_1 + (m_{21} + m_{22}\rho_1)}. \tag{6.12}$$

m_{11}, m_{12}, m_{21} and m_{22} are elements in a 2×2 matrix M written as:

$$M = \begin{bmatrix} m_{11} & m_{12} \\ m_{12} & m_{22} \end{bmatrix} = \begin{bmatrix} \cos\beta_2 & -\frac{i}{\rho_2}\sin\beta_2 \\ -i\rho_2 \sin\beta_2 & \cos\beta_2 \end{bmatrix}, \tag{6.13}$$

where β is equals

$$\frac{2\pi}{\lambda_i} n_2 h \cos(\theta_2). \tag{6.14}$$

For the wavelength of the AR coating to be centred at 10.59 μm, the ZnS layer must be 1.18 μm thick, since the refractive index of ZnS at 10.59 μm is 2.2432 [26]. The resulting calculated absolute transmittance through the layer at normal incidence is shown in figure 6.30.

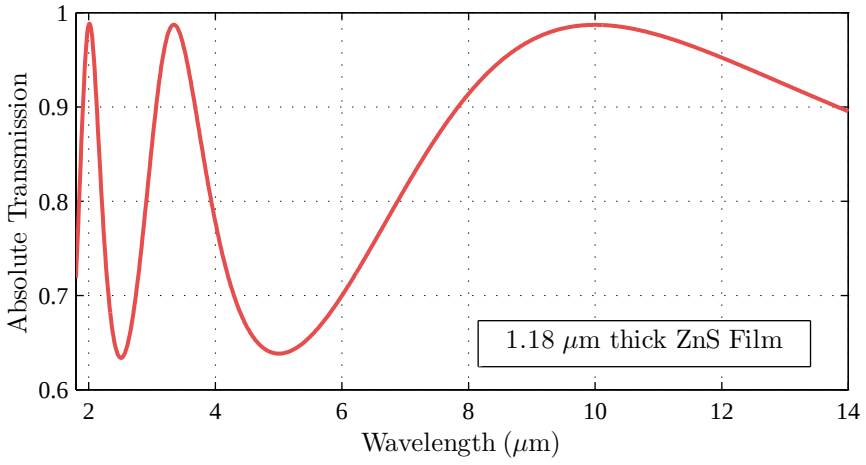


Figure 6.30: The *calculated* transmission through the layer as a function of incident radiation wavelength.

The transmission through the layer as a function of incidence angle θ is plotted in figure 6.31. The wavelength was fixed at 10.59 μm. As can be seen, below 50^o, the reflection is negligible. Thus the structure has a large acceptance angle which can accommodate f-numbers less than 1.

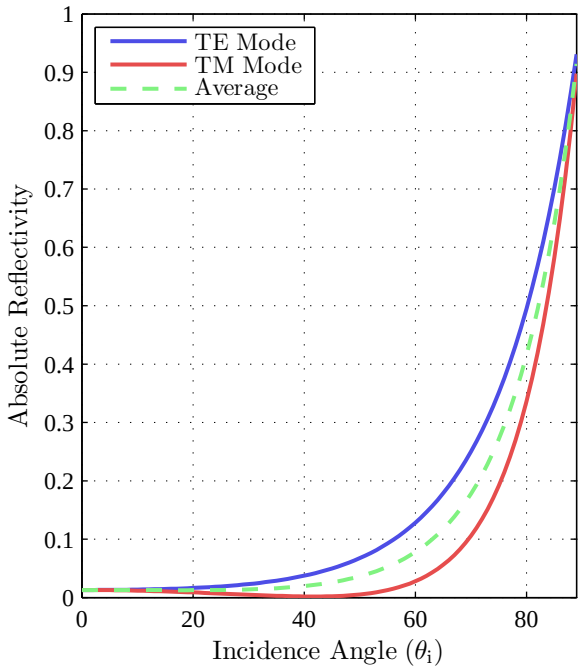


Figure 6.31: The reflectivity through the stratified layer as a function of angle of incidence. The wavelength was fixed at 10.59 μm. The average reflectivity was computed using $(R_{TE} + R_{TM})/2$

6.11.2 Double Sided AR Coated Intrinsic Germanium

ZnS coatings were deposited on intrinsic germanium samples in a thermal evaporator that has a rotating sample holder. The evaporation process is fairly straight forward and good thickness homogeniety was obtained. This can be inferred from figure 6.32. ZnS was deposited at rates below 1 nm/s to produce coatings that are more robust [1].

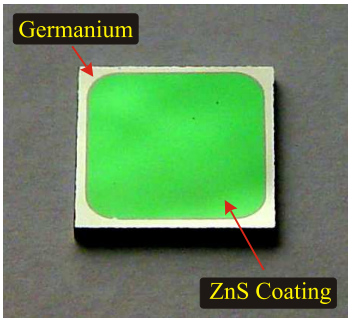


Figure 6.32: A piece of intrinsic germanium coated with 1.18 μm of ZnS.

The transmission through the sample is plotted in figure 6.33. Using just a single layer of ZnS, up to 99.5% transmission can be achieved at λ =10.59 μm. Without AR coating, the reflection losses from both surfaces of a germanium sample is $\approx$ 52%.

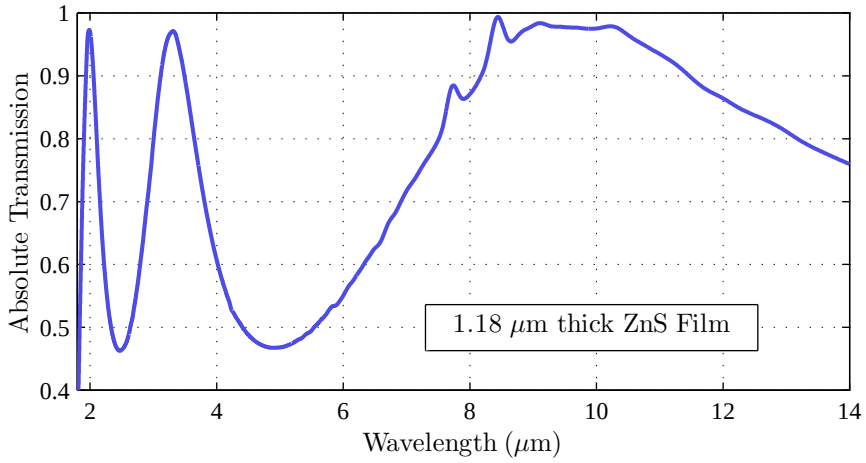


Figure 6.33: The transmittance through an AR coated germanium wafer as a function of wavelength, measured using a FTIR.

Although good quality ZnS AR coatings can be deposited once a routine is established, only one out of the seven modulators produced was coated due to time

limitation.

The use of ZnS AR coating that doubles as a passivation layer is a good step to be incorporated into the fabrication of future modulators for characterization purposes, since back reflection can be minimized and low SRV can be preserved.

6.12 Modulator Fabrication

The entire procedure to fabricate the modulator is illustrated in figure 6.34. Steps 1 to 12 have been described in the previous sections thus they will not be elaborated here. The core structure is completed at step 9, and an example of this structure is shown in figure 6.35. After step 12, the remaining procedure to complete the modulator is to attach the copper heat sinks and to provide the electrical contacts. Heat sinks are vital to ensure that the core does not heat up, since various secondary effects like the increase in attenuation due to thermally generated free carriers (figure 5.15) and the change in the intensity of transmitted light caused by the shift in interference fringes within the modulator can interfere with the optical characterization of the modulator [21]. For this reason, the thermal mass of the heat sinks is at least 50 times that of the core. To avoid the difficulty of soldering large wires directly to the core of the modulator (to support 2.6 A of current flow during forward bias) they were connected to the heat sinks instead (figure 6.36). For this design to work, indium foils must be used to fill the gaps between the flat surface of the modulator and the relatively rough surface of the heat sink to ensure that good electrical and thermal contacts can be established. Indium is used because it is soft enough to be deformed, has good thermal and electrical conductivity and does not need to be cured at elevated temperatures unlike conductive epoxies. The entire procedure to produce the modulator is thus complete.

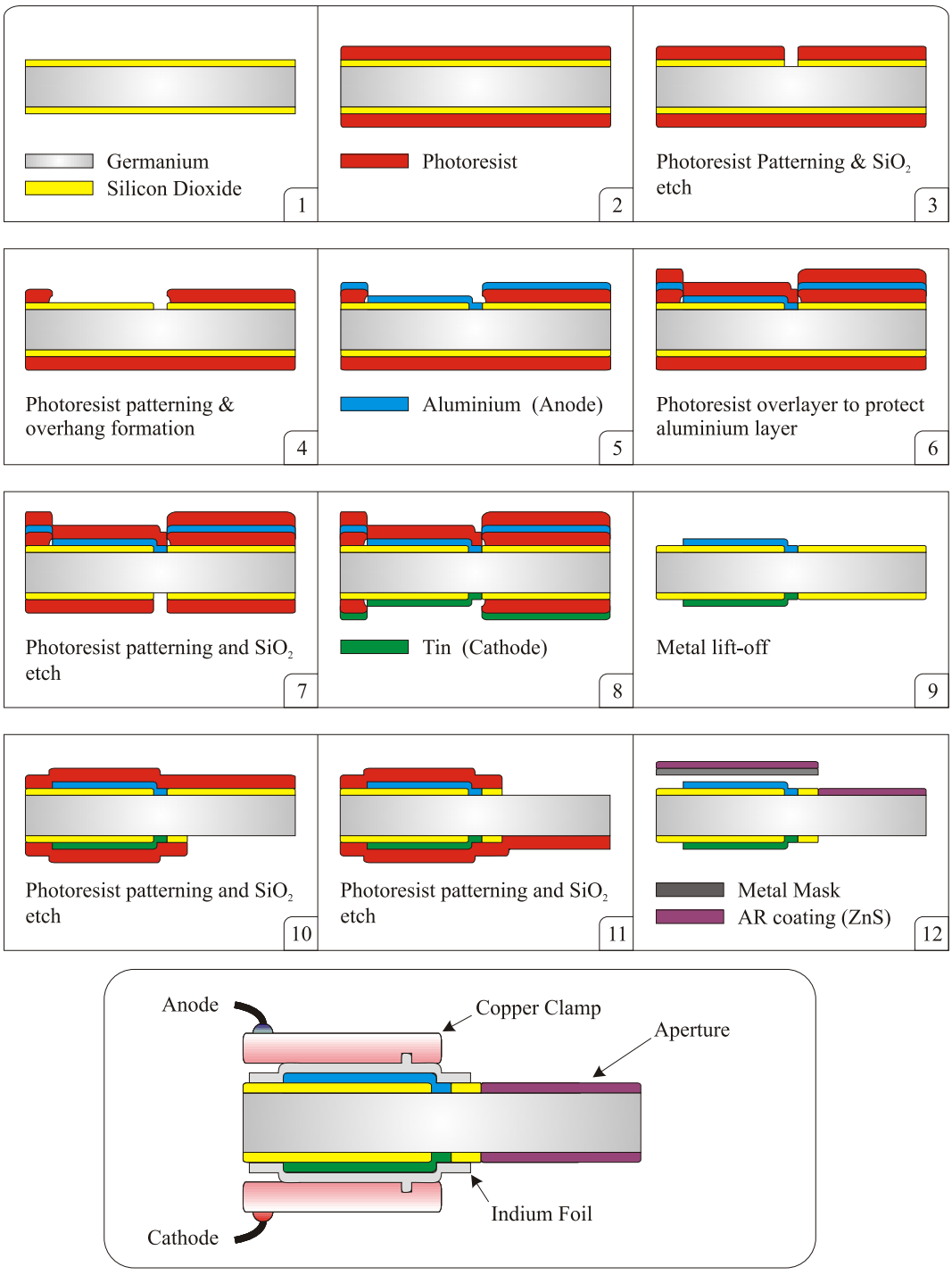
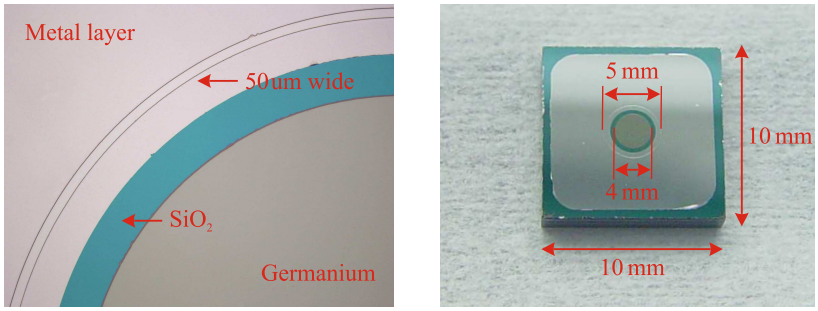


Figure 6.34: An outline of important steps required to produce a complete modulator.



(a) (b)

Figure 6.35: The core modulator structure. In (a) the metal layer is deposited on top of the dielectric layer, which has a narrow stripe of annular trench etched all the way to the surface of the sample. This allowes the metal layer to contact the germanium within that stripe, which then acts as the electrode of the modulator. Note also that the metal layer actually extends into the aperture, effectively reducing its diameter. This was done to allow some tolerance to device fabrication. (b) The dielectric layer also intrudes into the aperture, reducing its size from 5 mm to 4 mm. This was done to prevent the indium foil from contacting the aperture when it is clamped between the copper plate and the modulator (see figure 6.34).

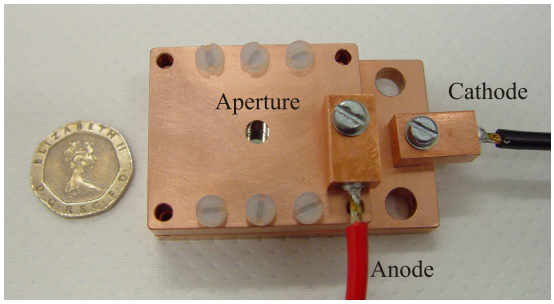


Figure 6.36: A modulator with copper heat sinks that double as electrical contacts.

Bibliography

[1] A. A. J. Al-Douri and O. S. Heavens. The role of rate of deposition and substrate temperature on the structure and loss of zinc sulphide and magnesium fluoride films. *Journal of Applied Physics D*, 16:927–935, 1983.

[2] L. D. Armstrong, C. L. Carlson, and M. Bentivegna. P-n-p transistors using high-emitter-efficiency alloy materials. *RCA Review*, 17:37–45, 1956.

[3] H. H. Berger. Models for contacts to planar devices. *Solid-State Electronics*, 15(2):145–158, 1972.

[4] M. Born and E. Wolf. *Principles of Optics*. Pergamon Press, 3^{ed} edition, 1965.

[5] A. Bratschun. The application of rapid thermal processing technology to the manufacture of integrated circuits—an overview. *Journal of Electronic Materials*, 28(12):1328–1332, 1999.

[6] C. Canali, J. W. Mayer, and G. Ottaviani. Solid-phase transport and epitaxial growth of ge and si. *Applied Physics Letters*, 25(1):3–5, 1974.

[7] Y. L. Chao, S. Prussin, and J. C. S. Woo. Preamorphization implantation-assisted boron activation in bulk germanium and germanium-on-insulator. *Applied Physics Letters*, 87(142102):1–3, 2005.

[8] C. O. Chui, K. Gopalakrishnan, P. B. Griffin, J. D. Plummer, and K. C. Saraswat. Activation and diffusion studies of implanted n and p dopants in germanium. *Applied Physics Letters*, 83(6):3275–3277, 2003.

[9] C. O. Chui, L. Kulig, J. Morgan, and Wilman Tsai. Germanium n-type shallow junction activation dependences. *Applied Physics Letters*, 87(091909):1–3, 2005.

[10] W. Cunningham, A. Gouldwell, G. Lamb, P. Roy, J. Scott, K. Mathieson, R. Bates, K. M. Smith, R. Cusco, I. M. Watson, M. Glaser, and M. Rahman. Probing bulk and surface damage in widegap semiconductors. *Journal of Applied Physics D*, 34:2748–2753, 2001.

[11] D. B. Cuttriss. Relation between surface concentration and average conductivity in diffused layers in germanium. *Bell System Technical Journal*, 40:509, 1961.

[12] B. Depuydt, A. Theuwais, and I. Romandic. Germanium: From the first application of czochralski crystal growth to large diameter dislocation-free wafers. *Materials Science in Semiconductor Processing*, 9:437–443, 2006.

[13] W. C. Dunlap. *An Introduction to Semiconductors*. John Wiley & Sons, 1st edition, 1960.

[14] P. D. Fairley. *Novel Solid State Modulator for the Infrared: The Germanium Chopper*. PhD thesis, University of Southampton, 2000.

[15] P. D. Fairley and H. N. Rutt. A novel gas sensing concentration monitor using a solid state modulator. *Sensors and Actuators*, 75:192–196, 2001.

[16] H. Geng. *Semiconductor Manufacturing Handbook*. McGraw-Hill, 2005.

[17] L. Holland. *Vacuum Deposition of Thin Films*. Chapman and Hall, 1st edition, 1963.

[18] J. D. Lee. *Concise Inorganic Chemistry*. Chapman and Hall, London, 4th edition, 1991.

[19] M. Lijadi, F. Pardo, N. Bardou, and J. L. Pelouard. Floating contact transmission line modelling: An improved method for ohmic contact resistance measurement. *Solid State Electronics*, 49:1655–1661, 2005.

[20] H. A. Macleod. *Thin-Film Optical Filters*. IoP, 3rd edition edition, 2001.

[21] C. D. Mansfield and H. N. Rutt. Evaluation of multiple beam interference effects in infrared gas spectroscopy. *Measurements Science Technology*, 10:206–210, 1999.

[22] C. W. Mueller and N. H. Ditrick. Uniform planar alloy junctions for germanium transistors. *RCA Review*, 17:45–56, 1956.

[23] G. Ottaviani, C. Canali, and G. Majni. Some aspects of ge epitaxial growth by solid solution. *Journal of Applied Physics*, 47(2):627–630, 1976.

[24] G. Ottaviani, V. Marrello, J. W. Mayer, M. A. Nicolet, and J. M. Caywood. Formation of injecting and blocking contacts on high-resistivity germanium. *Applied Physics Letters*, 20(8):323–325, 1972.

[25] J. P. Ponpon. Evolution with time the surface properties of high-purity germanium. *Nuclear Instruments and Methods in Physics Research A*, 457:262–265, 2001.

[26] R. F. Potter. Semiconductors. In E. D. Palik, editor, *Handbook of Optical Constants of Solids II*. Academic Press, 1991.

[27] F. E. Roberts. Impurity concentrations in the regrown region of the in-ge alloyed junctions. *Solid-State Electronics*, 2(1):8–13, 1961.

[28] A. Satta, E Simoen, T. Janssens, T. Clarysse, B. D. Jaeger, A. Benedetti, I. Hoflijk, B. Briks, M. Meuris, and W. Vandervorst. Shallow junction ion implantation in ge and associated defect control. *Journal of the Electrochemical Society*, 153(3):G229–G233, 2006.

[29] D. K. Schroder. *Semiconductor Material and Device Characterization*. John Wiley & Sons, New Jersey, 3rd edition, 2006.

[30] D. Sigurd and W. van der Weg. Solid-phase transport and epitaxial growth of ge and si. *Applied Physics letters*, 25(1):3–5, 1974.

[31] E. Simoen, A. Satta, M. Meuris, T. Janssens, T. Clarysse, A. Benedetti, C. Demeurisse, B. Brijs, I. Hoflijk, W. Vandervorst, and C. Claeys. Defect removal, dopant diffusion and activation issues in ion-implanted shallow junctions fabricated in crystalline germanium substrates. *Solid State Phenomena*, 108-109:691–696, 2005.

[32] W. G. Spitzer, F. A. Trumbore, and R. A. Logan. Properties of heavily doped n-type germanium. *Journal of Applied Physics*, 32(10):1822–1830, 1961.

[33] Y. S. Suh, M. S. Carroll, R. A. Levy, G. Bisognin, and D. D. Salvador. Implantation and activation of high concentrations of boron in germanium. *IEEE Transactions of ELectron Devices*, 52(11):2416–2421, 2005.

[34] S. M. Sze. *Semiconductor Devices*. John Wiley and Sons, 2nd edition, 2002.

[35] S. M. Sze and J. C. Irvin. Resistivity, mobility and impurity levels in gaas, ge and si at 300 k. *Solid-State Electronics*, 11:599–602, 1968.

[36] S. M. Sze and K. K. Ng. *Physics of Semiconductor Devices*. John Wiley & Sons, New Jersey, 3rd edition, 2007.

[37] F. A. Trumbore. Solid solubilities of impurity elements in germanium and silicon. *The Bell System Technical Journal*, 39:205–233, 1960.

[38] S. Uppal, A. F. W. Willoughby, J. M. Bonar, N. E. B. Cowern, T. Grasby, R. J. H. Morris, and M. G. Dowsett. Diffusion of boron in germanium at $800 - 900^o$c. *Journal of Applied Physics*, 96(3):1376–1380, 2004.

[39] K. R. Williams, K. Gupta, and Matthew Wasilik. Etch rates for micromachining processing—part ii. *Journal of Microelectromechanical Systems*, 12(6):761–778, 2003.

[40] P. V. Zant. *Microchip Fabrication*. McGraw Hill, 5th edition, 2004.

[41] J. Ziegler. Particle interactions with matter. *http://www.srim.org/*, page (accessed 15 September 2007), 2007.

Chapter 7

Experimental Results

In this chapter, the experimental techniques used to determine the performance of the modulator are discussed. Results on the attainable depth of modulation, the evolution of excess carriers during forward bias, the effective lifetime of carriers, and the usefulness of reverse voltage on switch off to reduce the recovery time are also presented. These results are compared with the predicted behaviour of the default simulation model described in chapter 5. The chapter ends with an analysis of the performance of the modulator between 77 and 340 K to gauge the feasibility of its use within the standard military temperature range of 300 ± 50 K.

7.1 Modulators Produced

Throughout the project, various prototypes of the modulator were constructed. However, only the last batch of devices made were capable of achieving a current flow of 2.6 A without dissipating more than 3 W of power. Seven such modulators with different physical parameters were produced, and their structural properties are listed in table 7.1. These were made to enable the performance of the actual modulators to be compared with simulation results. Unfortunately, the inhomogeneity in infrared absorption across the aperture of the modulators prevented the direct correlation between the depth of modulation and the change in the physical parameters of the modulator.

7.2 The Electrical Performance of the Modulator

Achieving good current modulation at low power dissipation poses one of the biggest challenges in the fabrication of the modulator . If the maximum power dissipation is limited to 2 W, and 2.6 A is required to achieve $\approx 95\%$ IR attenuation according to ATLAS, this implies that the total resistance between all the electrical contacts must not

Sample	Doping Method	Wafer Thickness (mm)	Electrode Width (μm)	Aperture Diameter (mm)
MI-1-50-5	Implantation	1	50	5
MI-2-50-5	Implantation	2	50	5
MI-2-250-5	Implantation	2	250	5
MI2-50-3	Implantation	2	50	3
MSPE-2-50-5	SPE	2	50	5
MSPE2-50-5-noA*	SPE	2	50	5
MSPE2-50-3	SPE	2	50	3

Table 7.1: The final batch of modulators fabricated for characterization. The surface of the aperture of each of modulator is doped except for the sample marked with (*).

exceed 0.29 Ω. Early prototypes of the modulator failed to produce IV characteristics that are comparable to the curve predicted by ATLAS due to parasitic resistances at the contacts. Nevertheless, by optimizing the contact deposition procedure (section 6.10), the final batch of modulators produced reasonable IV characteristics. The results are shown in figure 7.1.

By comparing the IV curve of the modulators with the simulated curve, the effective contact resistance for each device can be assumed to be<0.6 Ω. Since the contact resistance of the electrodes are less than $10^{-2}\,\Omega$ (table 6.7), additional resistance could arise from the oxide layers on the surface of the copper plates used to conduct current to the electrodes of modulator (figure 6.34). (The plates were etched in HCl acid prior clamping to the modulator to remove any oxide present.) Better modulator packaging must be devised to reduce the parasitic resistances in future devices.

7.3 CO_2 Laser Probing

A continuous wave CO_2 laser operated at 10.59 μm was used to characterize the performance of the modulators. This method is advantageous because it allows the temporal behaviour of the modulator such as the rise and fall time of carrier density to be determined at a response speed only limited by the choice of detector used.

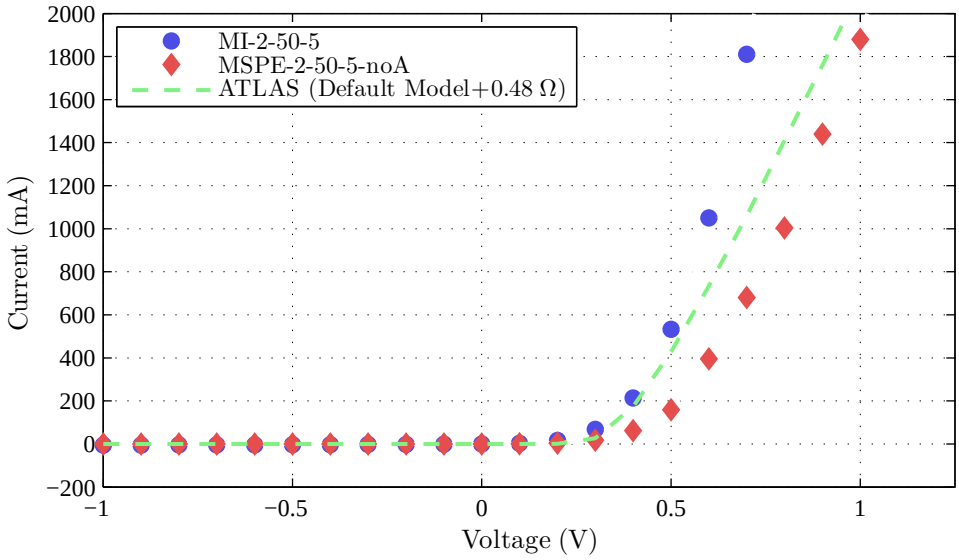


Figure 7.1: The IV characteristic of different modulators. To estimate the magnitude of parasitic resistance present in the circuit of the modulator, the IV curve of the default simulation model with 0.48 Ω of contact resistance was plotted alongside. The effective contact resistance of the fabricated modulators can be inferred to be less than 0.6 Ω.

7.3.1 Measurement Principle

The measurement principle can be described as follows: Excess carriers are injected into the modulator during forward bias and are probed by a CO_2 laser beam at $\lambda = 10.59\,\mu$m. If the transmittance is known, then density of carriers can be calculated using

$$T = T_0 \exp[-(\sigma_e \Delta n + \sigma_h \Delta p)]. \tag{7.1}$$

σ_e and Δn are respectively the absorption cross section and the density of electrons, while σ_h and Δp are the cross section and concentration of holes. (See section 2.3 for more details.)

7.3.2 Experimental Details

The experimental setup to analyze the performance of the modulator is shown in figure 7.2. The CO_2 laser has an output power between 4 and 6 W and a beam diameter of ≈ 6 mm. Two beam splitters are used to attenuate the beam power to <50 mW. This was done to prevent the thermal generation of carriers (through laser heating) which would add to the total infrared attenuation.

The probe beam is focussed at the aperture of the modulator using a ZnSe lens

with a focal length of 120 mm to a spot size of $\approx 100\,\mu$m. Strong focussing was not used to ensure that the probed volume loosely resembled that of a cylinder within the modulator for homogeneous probing. The transmitted beam is collected by a second ZnSe lens and refocussed again using a third ZnSe lens onto an uncooled HgCdZnTe photoconductor with a response time of less than 5 ns. Signals from the detector are amplified using a low noise pre-amplifier and sampled using an ADC card which then transfers the information to a computer.

The depth of modulation of the modulator is calibrated using a mechanical chopper, which gives a 100% depth of modulation when the beam is fully blocked. The transmittance obtained when the modulator is electrically on can then be scaled to this 100% signal to calculate the actual depth of modulation.

The temperature of the modulator is kept constant by the copper heat sinks. This is important to reduce etalon effects (i. e., the shifting of fringes) caused by the change in refractive index of the modulator at different temperatures [5].

7.3.3 Depth of Modulation

The depth of modulation map for two modulators are shown in figure 7.3. These were obtained by performing a 9×9 matrix scan on the 4 mm aperture. Up to a few thousand transients are averaged at each position to reduce the noise level. Data between each point was fitted using MATLAB to produce the map.

Inhomogeneous absorption were measured across the aperture of both modulators. This is caused by uneven current flow between the annular electrodes, where regions that conduct the majority of the current flow coincide with zones that show greater depth of modulation. This effect could be caused by uneven contact resistance around the annular electrode. The use of multiple constant current supply to power the modulator should solve this problem.

7.3.4 Transient Response

The transient behaviour of the modulator compares well with the simulated behaviour of the default modulator structure (figure 5.9). Less than 1 ms is needed to reach 80% depth of modulation for a two millimetre thick device (figure 7.4b). This implies that the current reaches steady state in a slightly faster time. If the simulation results hold, then a 4 mm thick modulator would need less than 2 ms to reach steady state at 2.6 A, the current required to achieve >95% attenuation at $10.59\,\mu$m.

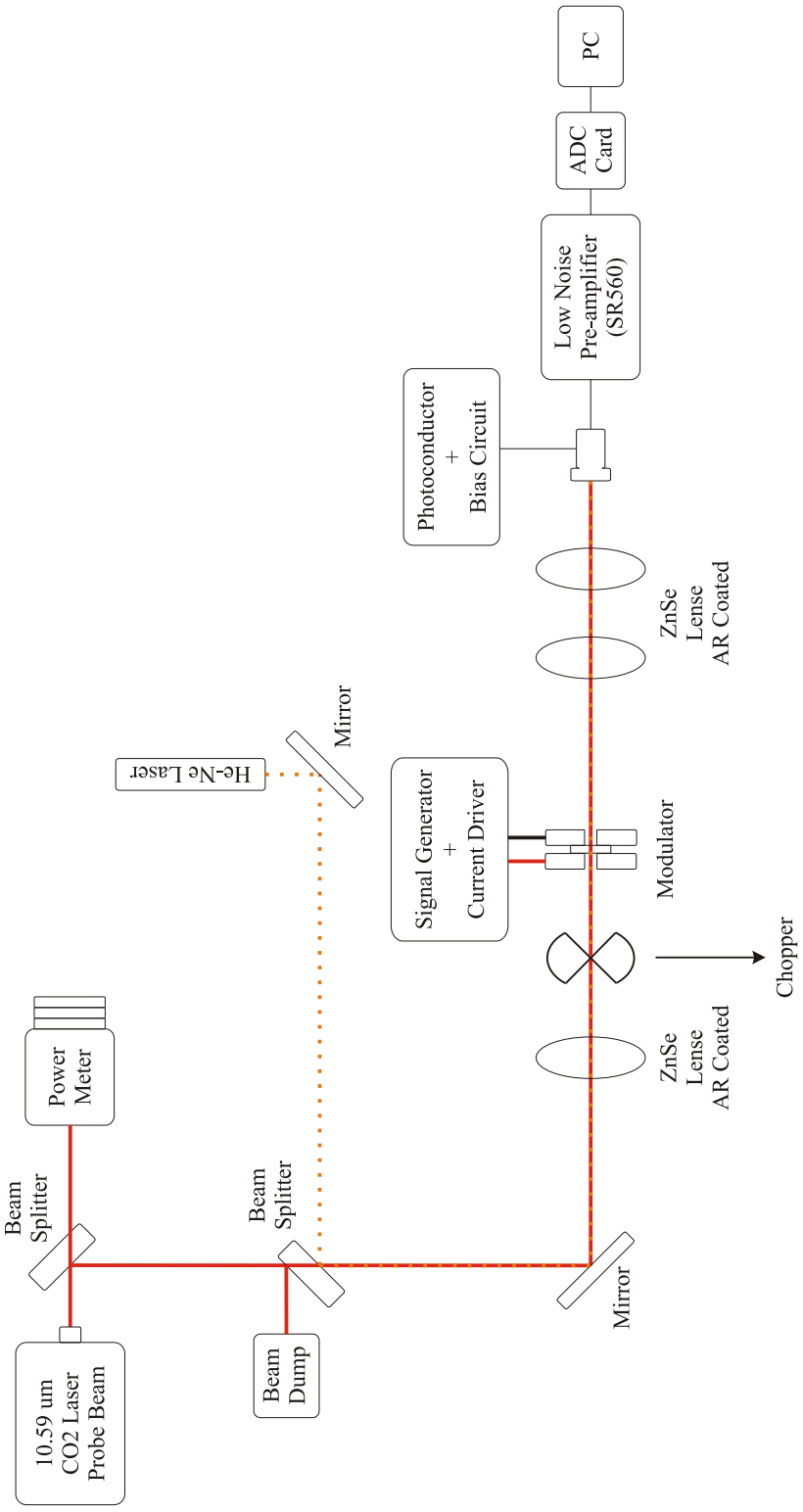


Figure 7.2: The experimental setup to analyze the optical properties of the modulator at 10.59 μm.

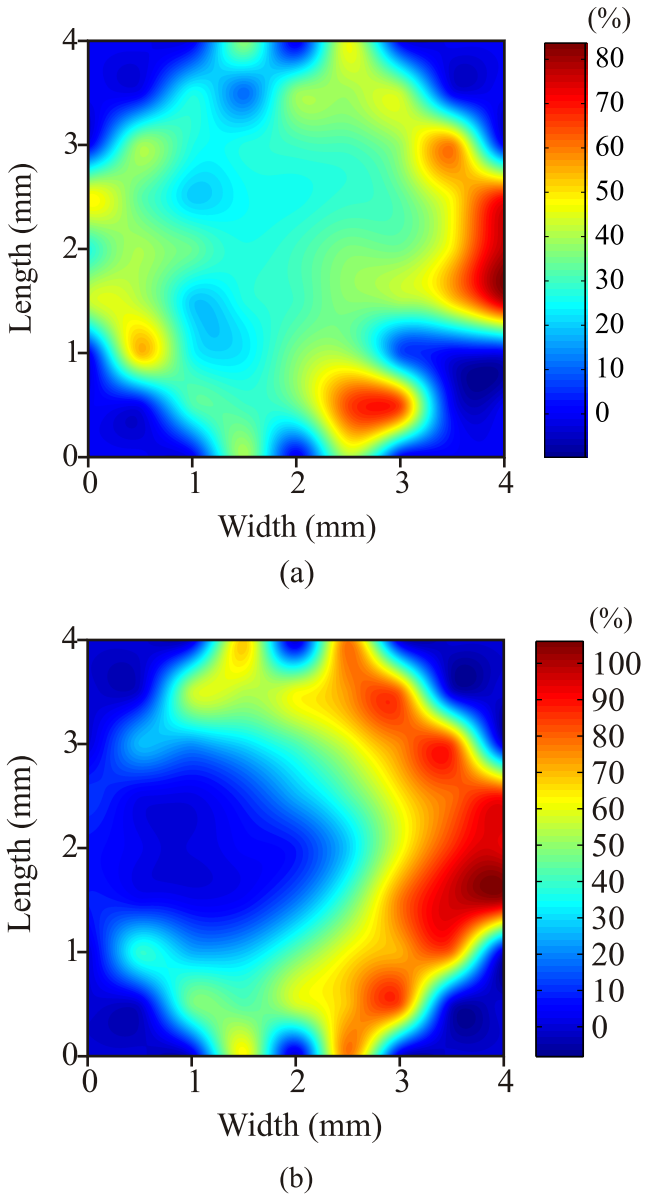


Figure 7.3: The total depth of modulation through the aperture of sample (a) MI-2-50-5 and (b) MSPE-2-50-5 operated at 2 A. Inhomogeneity in absorption across the aperture is cause by the uneven flow of current through the annular electrodes. More than 90% modulation can be obtained by electrically injecting carriers to the modulator.

The decay to zero however is slightly faster than predicted by ATLAS. This could be attributed to the higher rate of carrier recombination in the modulator, in which case SRH and other recombination mechanisms with similar dependency to radiative recombination would be responsible (section 7.4).

(The background noise with a characteristic frequency of 2.5 kHz is an electrical pickup that originates from the CO_2 laser. It is present even when the beam is blocked, and only disappears when the laser is turned off. We have not pinpointed the exact cause for this oscillation, though violent plasma oscillation, as a remote possibility, may somehow be coupled to the detector or its circuitry to cause the observe signal.)

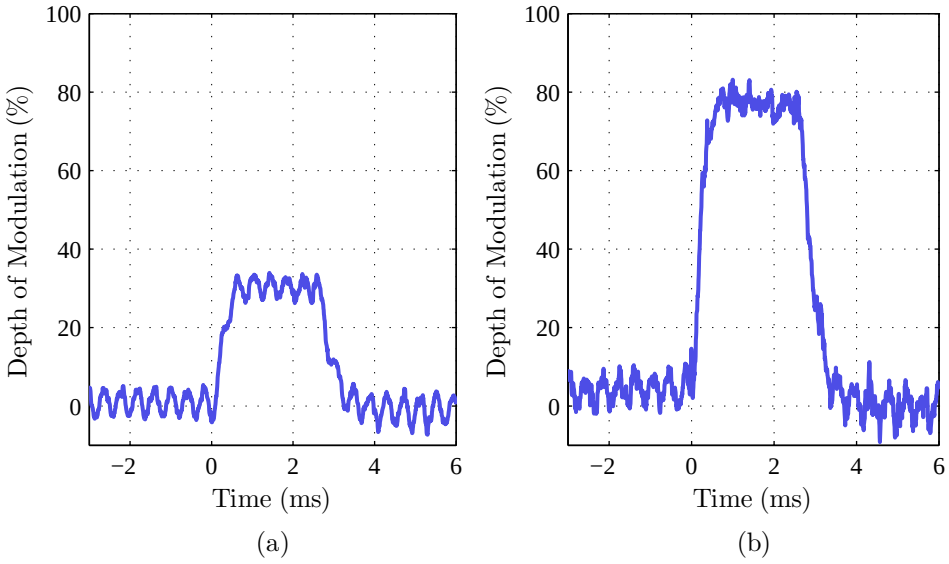


Figure 7.4: The depth of modulation and the transient behaviour of MI-2-50-5 operated at 2 A. The signals were sampled at (a) the middle of the aperture and (b) the edge of the aperture.

7.3.5 Device Recovery through Reverse Voltage on Switch Off

Reverse voltage on switch off was performed to speed the recovery time of the modulator. At 5 V reverse voltage, the decay in carrier density, inferred from the decrease in depth of modulation, is hardly noticeable. This conclusion is similar to results produced by ATLAS, which predicted a meagre decrease in the rate of carrier decay even at 10 V of reverse voltage (figure 5.27). This has been discussed in detail in chapter 5.

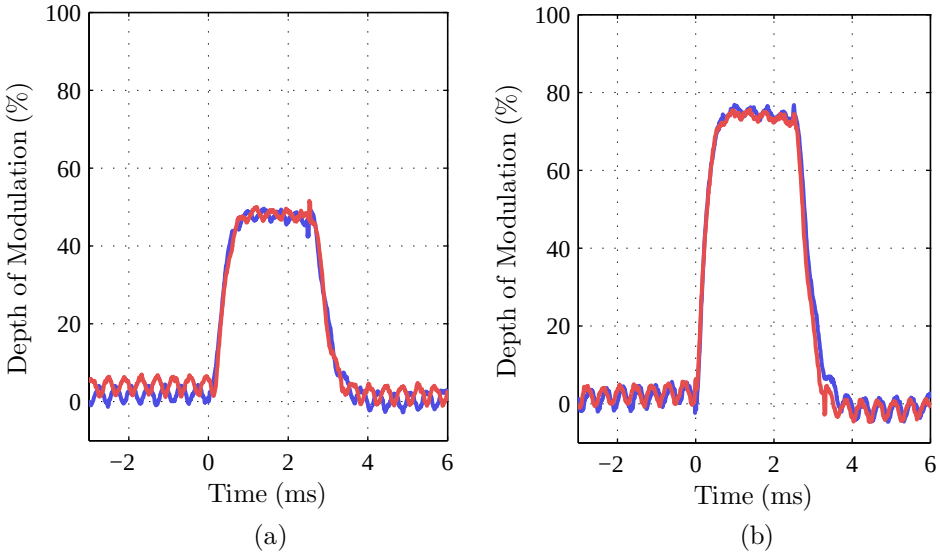


Figure 7.5: The recovery time for MSPE-2-50-5 with 0 and 5 V of reverse bias applied on switch off. The profiles were measured at (a) the middle of the aperture and (b) the edge of the aperture.

7.4 Carrier Lifetime Measurements

The determination of the effective lifetime of carriers is essential for the following reasons: (a) it is needed to accurately model the behaviour of the modulator using ATLAS, and (b) it provides insight into the performance of real devices. It is essential that carrier lifetime is characterized at the relevant carrier densities, since it may change by orders of magnitude. In addition, the measurement method used should not suffer from the dependence on parameters that cannot be determined accurately. Unfortunately, standard techniques to measure carrier lifetime including microwave reflectance, photoconductivity decay and surface photovoltage require the knowledge of the carrier mobility at the probed carrier density for the lifetime to be determined accurately [4, 7]. This makes them susceptible to errors when the density of carriers are evaluated. On the other hand, electrical lifetime measurement techniques do not permit the measurement of carrier lifetime as a function of carrier density [8]. Therefore, CO_2 laser probing was used to evaluate the lifetime of carriers because it is able to probe the density of carriers directly (provided the cross section is known), and is capable of resolving carrier lifetime at different carrier concentrations [4].

7.4.1 Experimental Results

The lifetime of carriers can be extracted from the decay in carrier density using the following equation:

$$\tau_{\text{eff}} = \frac{\Delta n}{\Delta R} = \frac{\Delta n}{-\dfrac{\mathrm{d}\Delta n}{\mathrm{d}t}}, \tag{7.2}$$

which gives the instantaneous lifetime of carriers at specific carrier density.

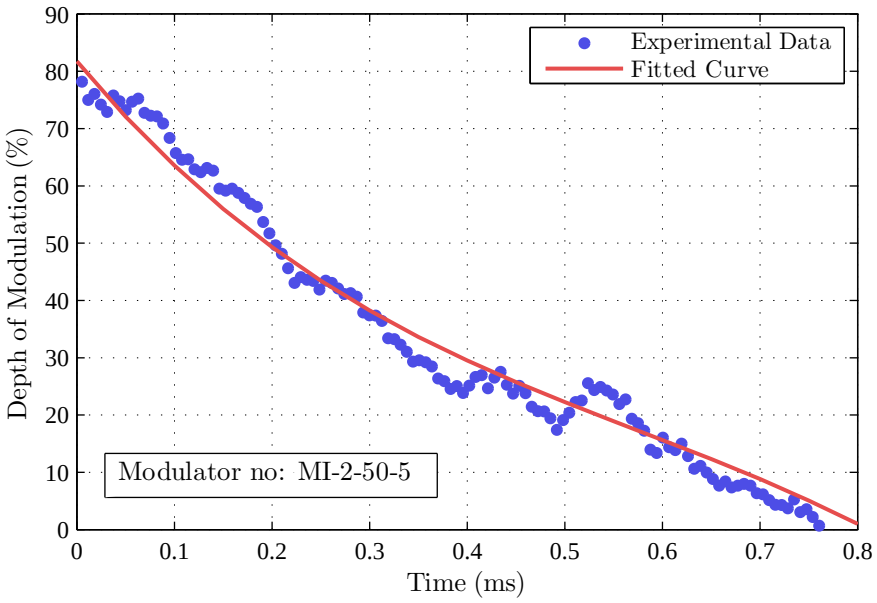


Figure 7.6: The reduction in depth of modulation with time for MI-2-50-5 measured from the edge of the aperture.

Data from figure 7.4b was used in the analysis of carrier lifetime, and the relevant part of the curve is re-plotted in figure 7.6. Because it is hard to compute the gradient using the existing data (figure 7.7), it was fitted with a curve generated using a third order polynomial equation in MATLAB. The gradient of this curve was then used in equation 7.2 to estimate the lifetime of carriers.

The resulting instantaneous carrier lifetime is plotted in figure 7.8. Three different injection regimes are resolvable: (a) at low carrier density, the effective lifetime of carriers is determined by surface and SRH recombination; (b) at moderate carrier densities where the concentration of carriers exceed the density of traps (i. e., SRH recombination), the lifetime increases. (c) the dip in lifetime at densities above $10^{14}\,\mathrm{cm}^{-3}$ is due to another recombination mechanism not related to auger recombination, since it only dominates at carrier densities above $10^{17}\,\mathrm{cm}^{-3}$ (figure 4.4). Landsberg attributed

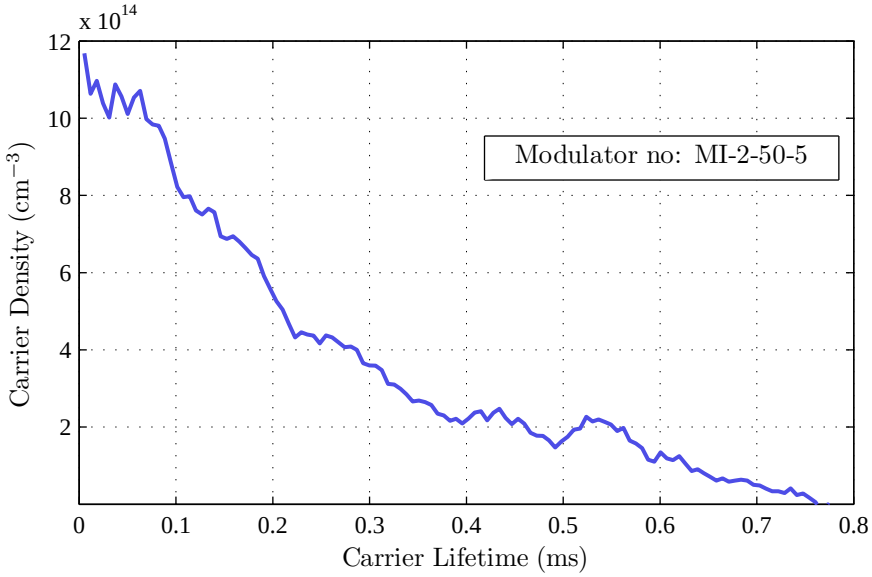


Figure 7.7: The decay of carrier density calculated from data plotted in figure 7.6. Equation 7.1 was used to estimate the concentration of carriers.

this effect to trap-assisted auger recombination, which obeys the same functional dependence as radiative recombination [3]. Therefore, its contribution can be lumped together with the radiative recombination coefficient β. However, in lifetime experiments conducted by Linnros using intrinsic and doped silicon that differ in impurity concentration by two orders of magnitude, the lifetime of carriers can only be fitted accurately using similar value of β. Thus it is unlikely that trap-assisted Auger recombination is responsible for dip in carrier lifetime for the intrinsic sample. (Incidentally, the value of β Linnros used is five times higher than the value quoted for silicon [7].) However, Linnros did not suggest an alternative recombination mechanism to corroborate the use of a second order recombination term (usually attributed to radiative recombination) to fit the lifetime of carriers and neither did he attributed it to radiative recombination. Irrespective of the actual mechanism, both Landsberg and Linnros had to include a recombination term that has the same functional dependence as radiative recombination to accurately fit the lifetime of carriers in silicon.

Similarly, a better fit to the instantaneous lifetime shown in figure 7.8 (though not very accurate) can only be obtained if the radiative recombination term is added. Arguments for including the radiative term to estimate the lifetime of carriers in indirect semiconductors like germanium are given in section 4.1.4.

The recombination parameters used to produce fit 2 are similar to the ones used in

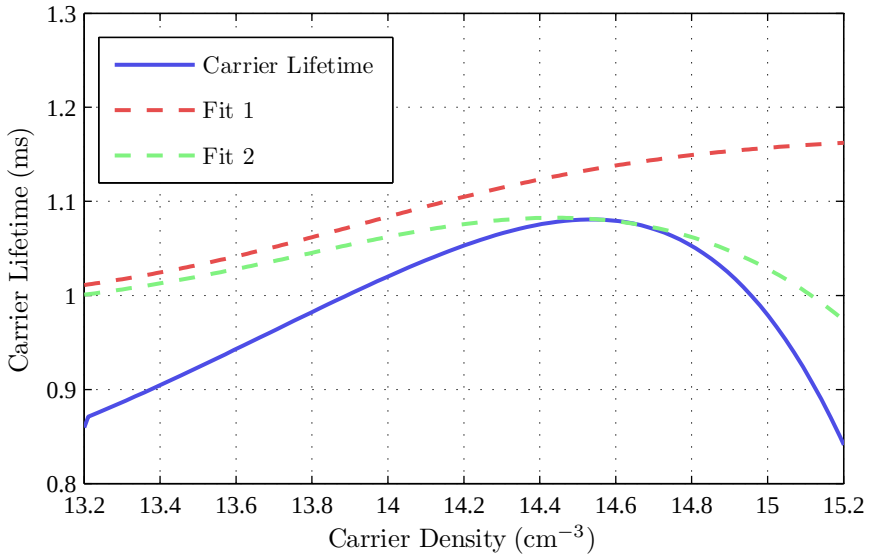


Figure 7.8: The instantaneous lifetime of carriers calculated from the fitted decay curve in figure 7.6. Fits 1 and 2 were produced using analytical equations presented in chapter 4. The recombination parameters used in both fits are: SRV=300 cm/s; $N_T = 3.65 \times 10^{12}$ cm^{-3}; and C_n and C_p of 10^{-31} cm6/s. The second fit includes the radiative recombination term that was calculated using a β of 5.2×10^{14} cm^3/s.

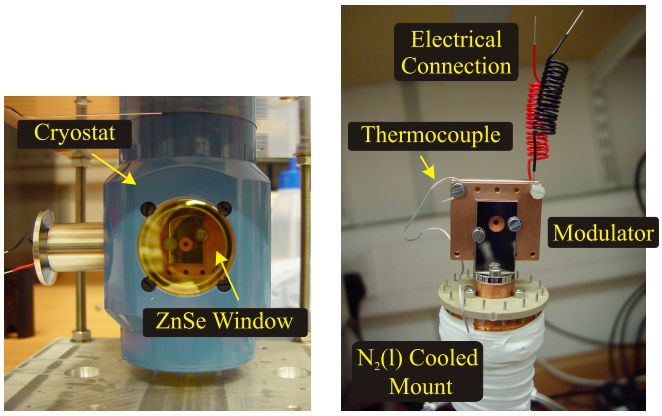
device simulation listed in table 5.1. Results from figure 7.8 show that lifetime parameters used are slightly optimistic. We could not produce better fits to the carrier lifetime using simple analytical equations, thus it is not possible to determine the recombination coefficients accurately. However, the severity this problem does not present a cause for concern because even devices with overly pessimistic recombination parameters are capable of achieving up to 98% depth of modulation according to ATLAS (figure 5.18). Results obtained from lifetime measurements shows that the modulator is not heavily contaminated by impurities, since the effective lifetime of carriers in regions where SRH recombination dominant is above 800 μs. In addition, radiative recombination (and/or any other equivalent recombination mechanisms) must be included to produce better fits to carrier lifetime. This substantiates the use of radiative recombination in modelling the behaviour of our devices (chapter 5).

7.5 The Absorption Cross Section as a Function of Temperature

For the modulator to replace the functionality of mechanical choppers in systems that use pyroelectric detectors, or if it is used for calibrating MCT based detectors/thermal imagers working at $8 - 14\,\mu$m, it should, ideally, modulate IR light consistently between $300 \pm 50\,$K. This means that the absorption cross section should not vary much within this range. Specifically, it is the absorption cross section for holes that matters because its magnitude is 16 times larger than that of electrons. Therefore, the theoretical estimates and the experimental data of the dependence of the hole absorption cross section (σ_h) with temperature between 150 and 340 K are presented in this section. These were performed to verify the usefulness of the modulator at temperatures other than 300 K.

7.5.1 Low Temperature Transmission Measurements

To conduct low temperature transmission measurements, the modulator is mounted on a liquid nitrogen cooled cryostat (Optistat DNV from Oxford Instruments) that can be evacuated to 10^{-6} mBar using a turbo pump. The section of the cryostat which houses the modulator is then positioned inside the sample compartment of a FTIR spectrometer for measurements to be performed.



(a) (b)

Figure 7.9: The cryostat used for low temperature transmission measurements. (a) A KF adaptor was attached to one of the viewing port provide the electrical feedthrough to the modulator and (b) the sample mount of the cryostat.

7.5.2 Transmission Through P-type Wafers

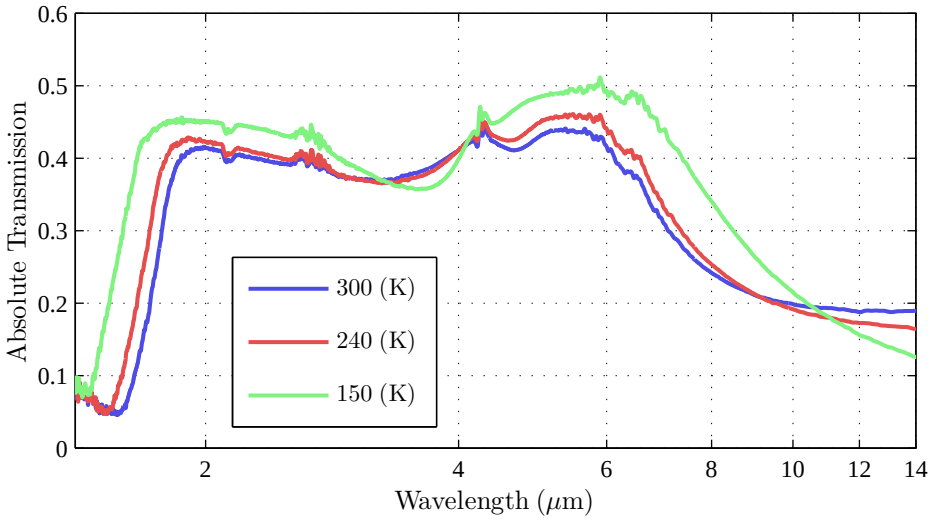


Figure 7.10: The transmission through a 250 μm thick wafer with $10^{16}\,\mathrm{cm}^{-3}$ average acceptor concentration.

A 250 μm thick p-type wafer with average acceptor concentration of $10^{16}\,\mathrm{cm}^{-3}$ was used for cross section measurements. The transmission through the wafer at three different temperatures is shown in figure 7.10. At 300 K, two distinct peaks located at 3.18 and 4.68 μm, and a broad absorption band between 8 and 14 μm can be seen. As the temperature drops, the 3.18 μm peak shifts to longer wavelengths while the opposite occurs for the 4.68 μm peak. Both peaks eventually superimpose at lower temperature, as can be seen in the curve measured at 150 K. The broad absorption peak shifts to longer wavelength when the temperature is reduced.

The aforementioned absorption profiles are caused by the excitation of electrons between the valence subbands (figure 7.11) [6]. Electrons can only be excited to lower energy subbands (energy increases in the negative direction in the valence band) if they are not fully occupied with electrons. The region depleted of electrons in the valence band is shown in pink in figure 7.11. As the temperature decrease, the region shrinks as electrons fill the lower energy bands. When this happens, all three transitions (A, B and C) occur at smaller momentum values. As a result, the maximum energy allowed by electrons to take part in transitions A and C is reduced, in accordance with the structure of the subbands. Transition B on the other hand requires more energy to take place, in agreement with results shown in figure 7.10.

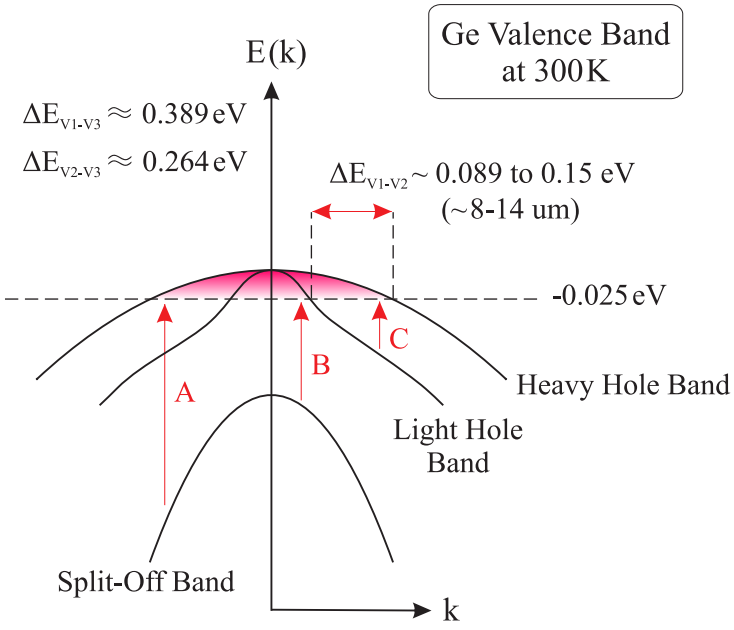


Figure 7.11: Germanium's valence band structure near Γ at 300 K (not to scale). Interband carrier transitions A and B give rise to IR absorptions centred at 3.18 μm and 4.68 μm respectively, while transition C results in the broad absorption between $8 - 14\,\mu$m

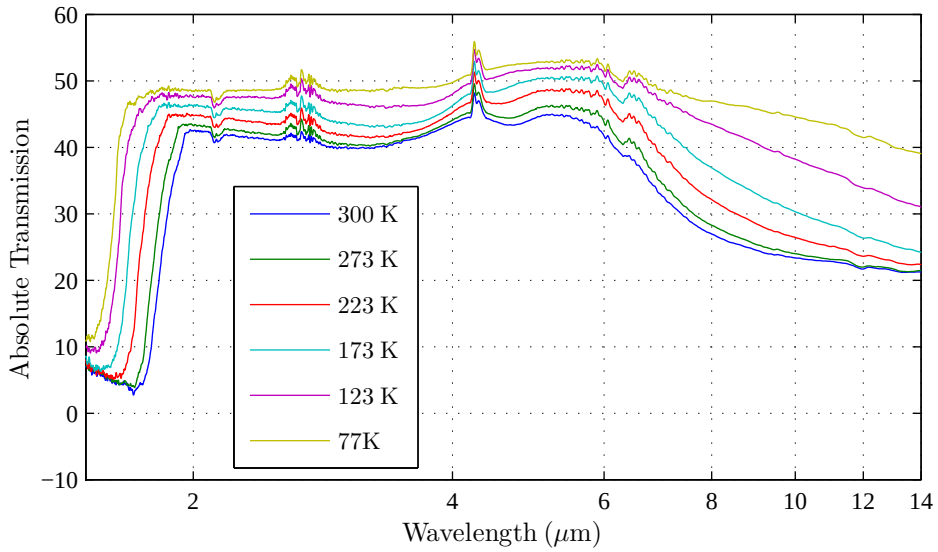


Figure 7.12: The transmission through MI-2-50-3 at 1.2 A of current.

A similar behaviour is observed for holes that are injected into the modulator. This is shown in figure 7.12. (The reduction in depth of modulation at lower temperatures is caused by the freezing of donor and acceptor ions in the 'n' and 'p' layer of the modulator, which reduces the injection efficiency of carrier and ultimately, the total number of carriers that flood the modulator. see figure 3.6)

7.5.3 Hole Cross Sections

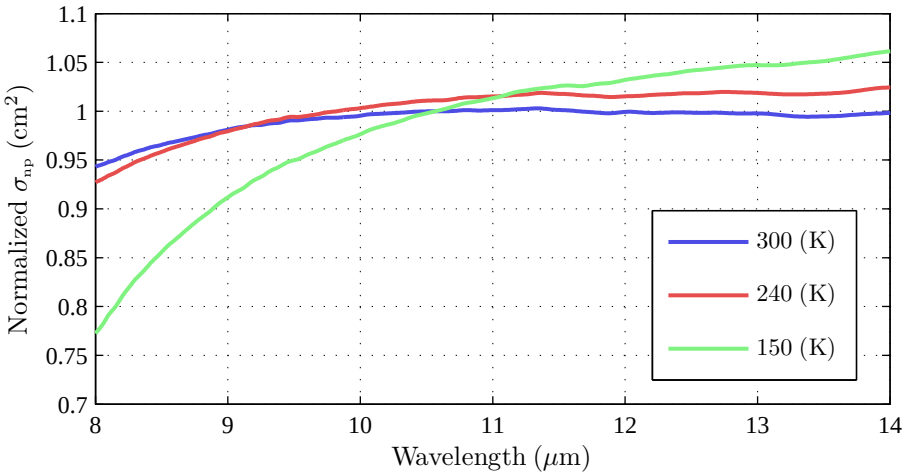


Figure 7.13: The normalized cross section calculated from the curves of figure 7.10.

The cross sections for holes normalized to 300 K and $\lambda = 10.59\,\mu$m are plotted as a function of wavelength and temperature in figure 7.13. Data from figure 7.10 and equation 7.1 were used to produce the cross sections. At 240 K, the cross section for wavelengths between 9 to 14 μm are slightly higher compared to that of 300 K, while the reverse is true between 8 to 9 μm. At 240 K, the difference in σ_p relative to 300 K between 8 and 14 μm is less than 3%. At lower temperatures however, the change in cross section is more drastic at both ends of the spectrum.

No experimental data is available for temperatures higher than 300 K. However, theoretical estimates on the cross section for holes can be extracted from band structure calculations [2]. Based on the work of Kane [2], the calculated hole cross sections for germanium as a function of temperature are shown in figure 7.14 [1]. When compared with data obtained experimentally, the general behaviour of σ_p with wavelength

and temperature is quite similar, except that the theory predicted bigger changes for σ_p. At shorter wavelengths, σ_p increases with temperature while the opposite is true at longer wavelengths. The change in σ_p between 8-14 μm at the highest calculated temperature did not exceed 10% when compared with the curve at 300 K. If the trend holds (i. e., theory predicting bigger changes) then the actual decrease in σ_p should be less, though it must be remembered that changes to σ_p affects the attenuation exponentially. Therefore, it is feasible to use the modulator within the temperature range of 300 ± 50 K.

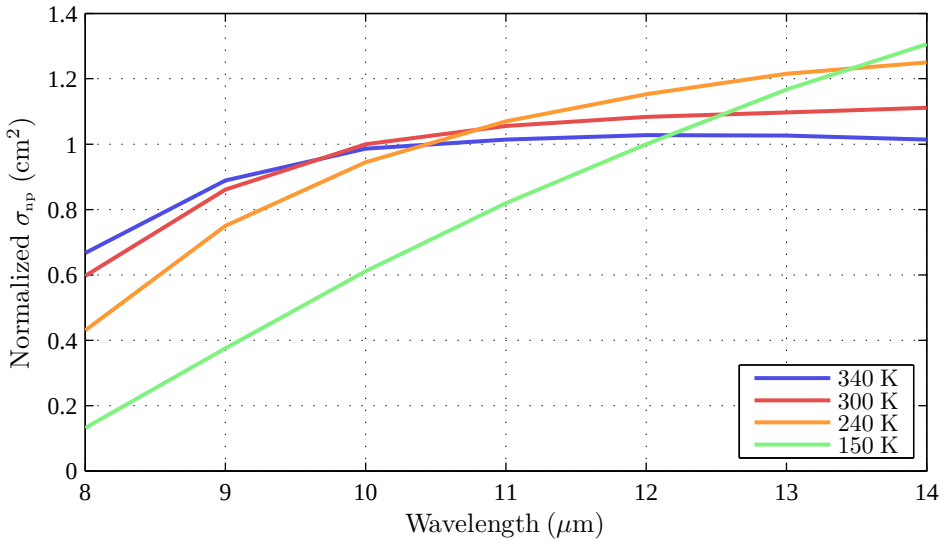


Figure 7.14: The calculated capture cross section normalized to 10.59 μm at 300 K [1].

Bibliography

[1] Private Communication. *The curves were obtained from H. N. Rutt who used the $k \cdot p$ method to calculate them.*

[2] E. O. Kane. Energy band structure in p-type germanium and silicon. *Journal of Physics and Chemistry of Solids*, 1:82–99, 1956.

[3] P. T. Landsberg. Trap-auger recombination in silicon of low carrier densities. *Applied Physics Letters*, 50:745, 1987.

[4] J. Linnros. Carrier lifetime measurements using free carrier absorption transients. i. principle and injkection dependence. *Journal of Applied Physics*, 84(1):275–282, 1998.

[5] C. D. Mansfield and H. N. Rutt. Evaluation of multiple beam interference effects in infrared gas spectroscopy. *Measurements Science Technology*, 10:206–210, 1999.

[6] J. I. Pankove. *Optical Processes in Semiconductors.* Dover Publications, New Jersey, 2^{nd} edition, 1975.

[7] D. K. Schroder. *Semiconductor Material and Device Characterization.* John Wiley & Sons, New Jersey, 3^{rd} edition, 2006.

[8] B. Tien and C. Hu. Determination of carrier lifetime from rectifier ramp recovery waveform. *IEEE Electron Device Letters*, 9(10):553–555, 1988.

Chapter 8

Conclusions and Future Work

8.1 Conclusions

The work involved in this project can be divided into three parts: device simulation, modulator fabrication and device characterization.

To ensure that appropriate physical models are invoked in ATLAS, all relevant analytical equations that govern the behaviour of carriers in germanium such as the energy gap, carrier mobility, and the recombination lifetime were reviewed in chapters 3 and 4. In chapter 5, the performance of the modulator predicted using ATLAS are presented. Up to 95% depth of modulation is possible using the default modulator specifications. Optimization to device structure like the use of thicker wafers, narrower electrodes, and higher doping concentrations in regions under the electrodes resulted in a factor of five improvement in the depth of modulation. In addition, the operating frequency of the modulator can be increased using pulse width modulation, though the effects are only significant for thicker samples. The entire procedure to produce the modulator using planar processing technology is covered in chapter 6. The first electrically operated solid state modulator for the $8-14\,\mu$m band was produced. In addition, alternative methods to dope the modulator were met with success, as pin diodes were successfully produced using solid phase epitaxial regrowth. Results from the bulk of experiments conducted are presented in chapter 7. Lifetime measurements showed that radiative combination must be included in estimating the lifetime of carriers. Free carrier absorption measurements using CO_2 laser probing has shown that up to 95% modulation can be achieved using electrically injected carriers. However, uneven current flow across the electrodes resulted in inhomogeneity in absorption across the aperture.

8.2 Future Work

Further work on all three areas of the project: device simulation, fabrication and characterization can be carried out.

In device simulation, most of the experimental data and physical models used date back to the early fifties and sixties. These data may not accurately describe the properties of ultra pure germanium wafers used in this project. Examples include the data for carrier mobility at different electric fields and doping densities, and the instantaneous lifetime of carriers at different concentrations (as highlighted in chapters 3, 4 and 7). Future simulation work should be carried out using more accurate experimental data (once they become available) to enable better predictions to the behaviour of the modulator. It must be stressed that no major changes to the properties of the modulator are expected.

Optimizing the use of pulse width modulation (PWM) to increase the operating frequency of the modulator should also be considered. At present, we have not found a suitable method to obtain the optimum pulse shape and repetition rate other than by trial and error. Though cumbersome, this procedure can be automated by writing a program in C that works in conjunction with ATLAS to obtain the optimum pulse signal via an iterative procedure. To achieve this, the program must be able to read the output of ATLAS after a simulation run (which is in ASCII format), analyze the current and voltage characteristics of the modulator, and subsequently generate a command file with updated biasing profiles (with upper and lower voltage and current limits set by the user) for ATLAS to process.

Finally, as discussed at the end of chapter 5, one way to increase the depth of modulation of diodes with apertures larger than 5 mm is to use a meshed electrode structure. This configuration enables carriers to be injected at multiple locations across the aperture. However, the drawback of this method is the reduced transmission through the modulator when it is electrically off—the main reason why this structure was not considered in this project. Nevertheless, it must be used to prevent unacceptable absorption inhomogeneities across large apertures. The effectiveness and feasibility of the mesh electrode structure can be evaluated using ATLAS.

In device fabrication, one vital issue that needs to be addressed is the reduction/elimination of parasitic resistance at the contacts—which is the cause for the uneven flow of current through the annular electrodes of the modulator. This problem

can be overcome by incorporating plasma etching in device fabrication to remove native oxides off metal contacts, and by optimizing the surface cleaning step of the germanium wafer. Failure to overcome this problem not only affects the performance of the modulator with annular electrodes, but also the future implementation of a functional modulator with meshed electrodes.

Optimizing the incorporation of donors (n-type dopants) in germanium using SPE would also benefit the fabrication of the modulator, since it can be done relatively cheaply and quickly. A way forward would be to replace antimony with arsenic, as literature have shown that higher peak doping density and fewer damages to the lattice can be achieved using the latter.

As for device characterization, it was within the original scope of this project to experimentally verify the validity of results predicted by ATLAS. Examples include the increase in depth of modulation when thicker substrates and narrower electrodes are used, and the increase in carrier lifetime when surface band bending is utilized. For this reason, seven modulators with different structural attributes were produced on the forth design iteration (table 7.1). Unfortunately, these devices could not be characterized reliably due to the uneven flow of current through the modulator, which resulted in inhomogeneous absorption across the aperture. As can be appreciated, the bulk of the time spent on this project were devoted to device fabrication (see the introduction of chapter 6). Because of time constraint, the fifth design iteration could not be pursued. Nevertheless, to ensure the validity of results produced by ATLAS, it is crucial that these experiments be performed. A good agreement between simulation and experimental results will support the validity of simulation work conducted for new structures (e. g., the meshed electrode structure). A quick way to mitigate the uneven flow of current through the modulators is to segment the electrodes and to supply current to each of the segments using separate current sources.

Appendix A

Simulation Code

```
#*********************************************
# Begin
#*********************************************
go atlas
# Runs ATLAS inside the Deckbuild environment
#*********************************************
# Section 1 : Mesh Specification
#*********************************************
mesh space.mult=1.5 cylindrical

x.mesh loc=0 spacing=40.0
x.mesh loc=2500 spacing=10.0
x.mesh loc=2750 spacing=10.0
x.mesh loc=3499 spacing=40.0
x.mesh loc=3500 spacing=0.001
x.mesh loc=3500 spacing=0.001
y.mesh loc=-0.005 spacing=0.001
y.mesh loc=-0.002 spacing=0.001
y.mesh loc=0.0 spacing=0.01
y.mesh loc=0.1 spacing=0.01
y.mesh loc=5.0 spacing=10.0
y.mesh loc=1000.0 spacing=40.0
y.mesh loc=1995.0 spacing=10.0
y.mesh loc=1999.9 spacing=0.01
y.mesh loc=2000.0 spacing=0.01
y.mesh loc=2000.002 spacing=0.001
y.mesh loc=2000.005 spacing=0.001

#*********************************************
# Section 2 : Structure Specification
#*********************************************
region num=1 material=Germanium

region num=2 material=oxide \
x.min=0.0 x.max=2500.0 \
y.min=-0.005 y.max = 0.0

region num=3 material=oxide \
```

```
x.min=0 x.max=2500.0 \
y.min=2000 y.max=2000.005

region num=4 material=oxide \
x.min=2750 x.max=3500 \
y.min=-0.005 y.max=0

region num=5 material=oxide \
x.min=2750.0 x.max=3500.0 \
y.min=2000.0 y.max=2000.005

region num=6 material=oxide \
x.min=3500 x.max=3500.005 \
y.min=0 y.max=2000

elec num=1 name=anode \
x.min=2500 x.max=2750 \
top
# anode

elec num=2 name=cathode \
x.min=2500 x.max=2750 \
bottom
# cathode

contact name=anode current
# Current boundary conditions

interface s.n=300 S.P=300 \
x.min=0 x.max=2500 \
y.min=-0.005 y.max=0

Interface s.n=300 s.p=300.0 \
x.min=0 x.max=2500 \
y.min=2000 y.max=2000.005

interface s.n=300 S.P=300 \
x.min=2500 x.max=3500 \
y.min=-0.005 y.max=0

interface s.n=300 S.P=300 \
x.min=2500 x.max=3500 \
y.min=2000 y.max=2000.005

interface s.n=300 S.P=300 \
x.min=3500 x.max=3500.005 \
y.min=0 y.max=2000

doping uniform conc=1e10 n.type
# Electrically active dopants found in the material 1e10/cm^3
```

```
# N and P type dopant profile via implantation
doping gaussian concentration=1e19 p.type \
dir=y peak=0.05 char=0.05 \
x.min=2500 x.max=2750

doping gaussian concentration=1e19 n.type \
dir=y peak=1999.95 char=0.05 \
x.min=2500.0 x.max=2750.0

#*******************************************
# Section 3 : Material Model Specification
#*******************************************
model temperature=300 print

MATERIAL MATERIAL=Germanium \
TAUP0=1.6 TAUN0=16 \
NSRHP=1e9 NSRHN=1e9 \
ETRAP=0.3323
# impurity concentration dependent srh for doped region

MATERIAL MATERIAL=Germanium \
COPT=5.2e-14
# radiative recombination

MATERIAL MATERIAL=Germanium \
AUGN=1.0e-31 AUGP=1.0e-31
# auger recombination coefficients

MOBILITY MATERIAL=Germanium ANALYTIC.N \
MU1N.CAUG=270.0 MU2N.CAUG=4300.0 \
NCRITN.CAUG=7e16 ALPHAN.CAUG=-1.66 BETAN.CAUG=-1.66 \
gamman.caug=-0.7 DELTAN.CAUG=0.636
# carrier mobility data, hand fitted and modelled using CONMOB

MOBILITY MATERIAL=Germanium ANALYTIC.P \
MU1P.CAUG=150.0 MU2P.CAUG=1800.0 \
NCRITP.CAUG=1.5e17 ALPHAP.CAUG=-2.33 BETAP.CAUG=-2.33 \
GAMMAP.CAUG=-1.7 DELTAP.CAUG=0.515
# carrier mobility data, hand fitted and modelled using CONMOB

MODELS OPTR CONSRH AUGER CONMOB ANALYTIC PRINT
OUTPUT U.RADIATIVE U.SRH U.AUGER

#*************************************
# Section 4 : Initial Solution
#*************************************
LOG OUTF=IV.log
METHOD NEWTON CARRIER=2.0 AUTONR
SOLVE INIT
SOLVE IANODE=1e-9 ISTEP=10 IMULT IFINAL=0.1 NAME=anode
SOLVE IANODE=0.2 ISTEP=0.4 IFINAL=2.6 NAME=anode \
```

```
OUTF=outslow00 MASTER
QUIT
```